DOCUMENT RESUME

ED 112 638 FL 006 855

AUTHOR Lehmann, Winfred P.; Stachowitz, Rolf A.
TITLE Development of German-English Machine Translation System. Final Technical Report.
INSTITUTION Texas Univ., Austin. Linguistics Research Center.
SPONS AGENCY Rome Air Development Center, Griffiss AFB, N.Y.
REPORT NO RADC-TR-75-2
PUB DATE Feb 75
NOTE 127p.; For related documents, see ED 062 888, ED 065 008, ED 105 716
AVAILABLE FROM National Technical Information Service, Springfield, Virginia 22161 (AD-A008 525/8GA, MF $2.25, PC $5.25)

EDRS PRICE MF-$0.76 HC-$6.97 Plus Postage
DESCRIPTORS Ambiguity; *Computational Linguistics; Computer Programs; *Contrastive Linguistics; Deep Structure; Descriptive Linguistics; English; Generative Grammar; *German; Idioms; *Machine Translation; *Programing; Pronouns; Sciences; Semantics; Sentence Structure; Structural Analysis; Surface Structure; Syntax; Technical Writing; Transformations (Language); Translation
IDENTIFIERS *Generative Semantics

ABSTRACT

This report describes work on a pilot system for a fully automatic, high-quality translation of German scientific and technical text into English and gives the results of an experiment designed to show the system's capability to produce quality mechanical translation. The areas considered were: (1) grammar formalism, mainly involving the addition of features to permit the generation of multiple translations for an imput sentence, (2) programming, consisting of completing the main components of the linguistic research system which is a generalized programming system for the mechanical translation of arbitrary languages, (3) linguistic description, concentrating on the coverage of the German surface syntax, a complete description of German and English standard grammar and syntactic choice grammar rules, and (4) lexicographic description. The characteristics of the system are: (1) the translation process is completely autonomous without human pre-analyses and/or pre-editing, (2) the output reflects quality features normally associated with a competent human translation, (3) the translation process is bidirectional, (4) the programs are language independent, and (5) the system is machine independent with respect to the programming language. (Author/TL)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED 112 638

RADC-TR-75-2
Final Technical Report
February 1975

DEVELOPMENT OF GERMAN-ENGLISH MACHINE TRANSLATION SYSTEM

The University of Texas at Austin

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

Approved for public release;
distribution unlimited.

FL 006865

Rome Air Development Center
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

APPROVED: Zbigniew L. Pankowicz

ZBIGNIEW L. PANKOWICZ
Technical Evaluator

APPROVED: Howard Davis

HOWARD DAVIS
Technical Director
Intel & Recon Division

FOR THE COMMANDER Carlo P. Crocetti

CARLO P. CROCETTI
Chief, Plans Office

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-75-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DEVELOPMENT OF GERMAN-ENGLISH MACHINE TRANSLATION SYSTEM		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report 6 July 1973 - 29 June 1974
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dr. Winfred P. Lehmann Dr. Rolf A. Stachowitz		8. CONTRACT OR GRANT NUMBER(s) F30602-74-C-0028
9. PERFORMING ORGANIZATION NAME AND ADDRESS The University of Texas at Austin Linguistics Research Center Austin, TX 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 31021F JO #ADHS0401
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRDT) Griffiss Air Force Base NY 13441		12. REPORT DATE February 1975
		13. NUMBER OF PAGES 122
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Zbigniew L. Pankowicz/IRDT
Griffiss AFB NY 13441
AC (315) 330-4231

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computational Linguistics
Machine Translation
Automated Lexicography
Computer Programming

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The report documents completion of Phase I out of three annual phases projected for development of the system for a fully automatic high quality translation of German S&T literature into English. The system under development has the following characteristics. Machine translation process is completely autonomous in the computer without pre-analysis and/or pre-editing of source language text. The output in target language reflects quality features normaliy associated with a competent human translation. Machine

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (cont'd)

translation process is reversible (German-English and English-German). Programs are language independent, thus allowing adaptation of the system to a bi-directional machine translation from any source language into any target language. The system is machine independent with respect to the programming language and, therefore, adaptable to any large scale IBM or CDC computing facility. On completion of the effort, as described in this report, the system has successfully passed a feasibility test for a fully automatic high quality translation, based on the German technical text.

ABSTRACT

Work on the completion of a pilot system for mechanical translation of German scientific and technical text into English is described. This report describes the effort performed in the area of Grammar Formalism, Programming, Linguistic Description and Lexicographic Description during the period from July 6, 1973 through June 30, 1974 under Contract F30602-74-C-0028. It supplements the work performed under contracts F30602-70-C-0118 and F30602-73-C-0192. This report also describes the results of an experiment performed during the reporting period which was designed to show the system's capability to produce quality mechanical translation.

Work on Programming consisted in completing the main components of the Linguistics Research System, a generalized progamming system for the mechanical translation of arbitrary languages.

Work on Grammar Formalism mainly involved the addition of features to permit the generation of multiple translations for an input sentence.

The linguistic work concentrated on the coverage of the German Surface Syntax, a complete description of German and English Standard Grammar and Syntactic Choice Grammar rules, as well as on the completion of lexical coverage for a text in the sciences prepared for the translation experiment.

INTRODUCTION

We stated in our proposal to this contract that any programming system designed to carry out fully automatic quality translation of random texts without pre- or post-editing has to take into account and provide solutions for the following linguistic problems confronting mechanical analysis of natural language:

a. discontinuous constructions, in particular verb prefix combinations,

b. idiomatic expressions,

c. empty words, which are not translated,

d. deleted terms with or without meaning change,

e. lexical collocations with internal variable slots,

f. category changes in translation,

g. sentence pattern changes in translation,

h. translation of the definite article in cases of inalienable property,

i. ambiguity resolution based on information within a sentence.

In addition to these points, we also required such a model to account for and solve the linguistic problems of anaphoric relations, in particular pronominal reference, the resolution of ambiguity based on information in the textual environment of a sentence, and preservation of input non-ambiguity. The Linguistics Research Center proposed to complete within the contract period a model for the mechanical translation from German-to-English. The first phase of the research effort had been sponsored under Contract number F30602-74-C-0028. In spite of the fact that the budget requirements were not completely satisfied, the Linguistics Research Center was able to construct within that contract period all major components of the proposed fully automatic quality translation system; this was possible on the one hand, because of the willingness of the linguists and programmers to work overtime without compensation, and on the other hand, through a most generous contribution of computer time by the University of Texas Computation Center. In a demonstration on June 21, 1974, the Linguistics Research Center successfully executed the fully automatic translation of a German text into English and proved the soundness and validity of the theoretical foundation of the Linguistics Research System and its viability. We were able to show that the system successfully solves the linguistic problems mentioned under points a) through c), e) through i), above; and to some degree, the problem stated under d).

The following sections describe the considerations leading to the experiment and its execution (Section I), the systems components and the linguistic problems encountered (Section II) and the considerations used to develop the lexical data base for the experiment (Section III). The appendix contains the German text and its translation.

TABLE OF CONTENTS

Abstract		i
Introduction		ii
I.	An Experiment in Mechanical Translation	I.1
1.0	Background	I.1
1.1	Purposes of the Experiment	I.3
1.2	Method	I.4
1.3	The Corpus	I.5
1.3.1	Lexicographic Description of RAK	I.5
1.3.2	Syntactic Description	I.5
1.4	Execution of Experiment	I.6
1.5	Results	I.8
II.	The Linguistics Research System (LRS)	II.1
2.0	Linguistic Basis	II.1
2.1	General Description of LRS Grammars	II.2
2.2	General Description of LRS Algorithms and their Interrelationship with the LRS Grammars	II.3
2.3	The Surface Component	II.4
2.3.1	The Dictionary Component	II.6
2.3.1.1	Dictionary Analysis	II.6
2.3.1.2	Dictionary Choice	II.10
2.3.2	The Word Component	II.12
2.3.2.1	Word Analysis	II.12
2.3.2.2	Word Choice	II.15
2.3.3	The Syntactic Analysis Component	II.18
2.3.3.1	Syntactic Analysis	II.18
2.3.3.2	Restrictions and Changes - Surface Grammar	II.20
2.3.3.3	Syntactic Choice	II.22

2.4 The Standard Components II.27

2.4.1 The lexical Collocation Component II.29
2.4.2 The Standard Analysis Component II.33
2.4.3 The Standard Choice Component II.35
2.4.4 The Transfer Components II.37

2.5 The Standard Production Components II.41

2.5.1 The File Entry Construction Component II.42

2.5.1.1 Interlingual Mapping II.42
2.5.1.2 Synthesis II.43
2.5.1.3 Standard Output Text II.44

2.5.2 Output Syntactic Choice - II.46

2.6 The Lexical Spellout Component II.48

III. Lexicographic Work Within the Experiment III.1

3.0 Introductory Remarks III.1
3.1 Methodology III.1

3.1.1 Provision of Translation Equivalents III.1
3.1.2 Coding of Rules III.3
3.1.3 Rule Coding for Ambiguous Items III.8

3.2 Lexical Features III.20

3.2.1 Noun Classification III.20
3.2.2 Adjective Classification III.26
3.2.3 Verb Classification III.27
3.2.4 Adverb Classification III.33

3.3 Compounds, Phrases, and Other Problems III.34

3.3.1 Noun Compounds III.34
3.3.2 Adjective Compounds III.35
3.3.3 Verbal Lexical Collocations III.36
3.3.4 Multi-word Prepositions III.40
3.3.5 Multi-word Adverbs III.41
3.3.6 Other Problems III.41

3.4 Concluding Remarks III.42

Conclusion III.43

Bibliography III.45

Appendix 1 A 1-11

TECHNICAL EVALUATION

"Development of German-English Machine Translation System"

The objective of this R&D effort consists in completion of Phase I out of three annual phases projected for development of an advanced operational capability in performing fully automatic high quality translation of German S&T literature into English.

The system under development is based on the current linguistic theory, commonly referred to as generative semantics. A generative semantics approach requires the assumption of underlying structures with a semantic, rather than purely syntactic, basis. From the viewpoint of translation this entails modification of surface structures to underlying structures in which semantic relations are taken into consideration. Past experience with second generation MT approaches has shown convincingly that high quality machine translation cannot be obtained by attempting to map input surface strings directly into output surface strings. The system represents the third generation approach, supplying the inherent deficiencies of second generation models, and incorporates the components of the current linguistic theory (surface component, deep component, transformational component and semantic component).

Translation process is based on mapping of the content of meaning through interlingual substitution of units of syntactic and semantic description. The proposed incorporation of contextual component will further enhance the output quality by permitting resolution of intra-sentential ambiguities with information derived from extra-sentential context.

The programs process input text with the system's grammars. They accept text in any form without human pre-analysis and/or pre-editing, such as indication of sentence boundaries. Text input can be prepared by any conventional means (punched cards, paper tape, OCR). The programs are independent of the content of grammar rules; they only expect the rules to be written in the required format. Consequently, the system can perform in both German-English and English-German translation mode without any changes in linguistic/lexical description of both languages and/or modification of programs. The grammar of any particular language can be described independently of considerations of its translatability into any other language. Adaptation of

the system to a bidirectional machine translation of other language pairs requires only the development of additional linguistic/lexical descriptions and no changes in the system's programs. High degree of machine independence with respect to the programming language allows to accommodate the system on any large scale IBM or CDC computer.

The system consists of grammars and programs. The grammars are componed of dictionary rules and non-dictionary rules. The latter describe the patterns occurring in sentences and parts of sentences. The format of the rules is convenient to learn; any kind or amount of syntactic/semantic information can be stated. Updating of dictionary rules and non-dictionary rules does not require any changes in programs, and can be performed by the user of the system independently without a continuous reliance on contractual assistance.

ZBIGNIEW S. PANKOWICZ
Technical Evaluator

SECTION I

1. An Experiment in Mechanical Translation

1.0 Background

From February 1970 through January 1971 the Linguistics Research Center performed a study with the purpose of ascertaining in some depth the then-current opinions of the linguistic and computational linguistic community on the Feasibility of Fully Automatic High Quality Translation. The report on this study has been published [1]. Among its conclusions was the statement that in spite of progress in linguistic analysis, linguistic research had dealt primarily with syntactic analysis of individual sentences and less with semantic problems and discourse analysis, thus making "current linguistic theory inadequate for machine translation". The recommendations made for the improvement of mechanical translation included the carrying out of studies on the incidence of lexical, syntactic, and semantic ambiguities in scientific texts, and on their resolution by means of information within a sentence (textual), in the environment of a sentence (contextual), and context external (factual) information. It was further recommended that explicit study be made of the kinds of information available to the user, which is necessary for the understanding of mechanically translated material. Such studies should seek to determine the amount of knowledge available from the actual text as well as the amount of world knowledge necessary for the understanding of individual sentences.

These recommendations, about which there was general consensus among the participants, were the outcome of the discussions in which, with varying degrees of conviction, opinions like the following had been voiced. Mechanical translation based on sentence-by-sentence translation would not be satisfactory.... Mechanical translation would have to take into account information provided by surrounding sentences (contextual information)..... Mechanical translation would have to have access to an encyclopedic component (factual knowledge)....Even human translation requires knowledge of the subject matter to be translated....For all these reasons, mechanical translation would only provide approximations of "acceptable", "good", or "quality" output*....In addition, texts might have to be pre-edited, output might have to be post-edited, or human assistance might even be required during intermediate processing stages, resulting in mechanical translation systems

*The term "fully automatic high quality translation" had fallen into disuse when Prof. Bar-Hillel, its originator, argued that this term could not really be defined as long as appropriate criteria had not been established.

ranging from human-aided mechanical translation to machine-aided human translation.

During these discussions it became clear that no comprehensive study had been performed to provide a factual foundation for these claims. The absence of such actual information was obviously due to the fact that no descriptive work had been performed to determine the amount of sentence internal and external information needed to disambiguate a particular sentence.

At the heart of these various opinions and proposals on mechanical translation systems, were actually two questions: 1) is or to what extent is human understanding of a text prerequisite to good translation; 2) to what extent could such human understanding be simulated and how costly would it be.

Was it indeed necessary that an MT system had to be equipped with an information retrieval system containing a knowledge of the world as some had stated? Or, would such a system need no factual information whatsoever, since the human reader could provide such information, as others had claimed? The aforementioned lack of empirical data left the issue undecided.

The statements that good mechanical translation requires the understanding of a text and facility with the subject matter discussed are actually based on the assumption that an MT system needs such information to disambiguate sentences, that is, to select one correct translation from a number of incorrect and/or correct translations.

In a paper [2], appended to that study, I described a mechanical translation system with a comprehension component which should be able to translate a test by exploiting, if required, information given in the text itself, i.e., in the textual environment for the disambiguation of a sentence. My conclusion, that such an MT system would provide good translations, was, of course, also conjecture, lacking actual factual confirmation.

The Linguistics Research Center was developing at that time a mechanical translation system, the Linguistics Research System (LRS) which is described in this and previous reports [3,4,5,6]. LRS contains a major portion of the components of the projected mechanical translation system described in that paper, the basic differences being that LRS is restricted to sentence-by-sentence translation, i.e., it does not contain a contextual and comprehension component. In spite of these limitations, we felt that LRS could be advantageously used to provide some answers to the questions which had been raised during the Feasibility Study. And we began the preparations for the mechanical translation of a popular scientific text, *Raketenantriebe (Rocket Propulsion)* by Joseph Stemmer [8].

1.1 Purposes of the Experiment

We began the experiment with different purposes in mind, hoping to obtain at least partial answers to the following questions.

1. How good a translation can be provided by a fully automatic system translating individual sentences only (that is, performing sentence-by-sentence translation without access to information in context)?

2. What kind of contextual information is needed for the disambiguation of individual sentences?

3. Is factual information required for the disambiguation of sentences, and if so, how much information and what kind?

4. Is knowledge of the subject matter required in order to disambiguate sentences?

We were certain that some kind of contextual component would be necessary, especially for the proper translation of pronouns as, for example, in German *wir beobachteten sie (we observed it/her/them)*. Our concern was mainly whether we would need a component of such complexity as described in "Requirements..." [2], or whether the information necessary for disambiguation could be provided inexpensively (according to one of the criteria by which MT is generally evaluated: quality, cost, and speed).

Apart from these questions, we were interested in obtaining mechanical confirmation of numerous hypotheses underlying our approach to MT. After all, no mechanical analysis or translation of texts had been performed at LRC since 1967. In that year we switched from an IBM 7040 to a CDC 6600. Before the conversion of the old LRC programs could be completed, a drastic cut in funding in late 1968 terminated for all practical purposes the work on the development of analysis and translation programs. Our work in such areas was only resumed in 1971 with the development of new algorithms designed to operate on grammars with complex symbols.

We were looking forward to the mechanical confirmation or falsification of a number of questions. Would our system be able to deal with sentences that contained non-words, like foreign citations, names? Would we be able to recognize potential sentence boundaries in order to obtain our goal of fully automatic translation of texts inputted by punch, OCR, or paper tape? Would it be necessary to change our lexicographic classification either by refining existing features or by adding completely new ones? Would our scheme to reduce the number of translations by preserving stylistic similarity of input and output, be successful? And numerous others. We were also interested in the occurrence of parasitic readings (interpretations provided by the algorithm during the various analysis phases but which are not part of the final correct analysis) in order to find means of restricting if not avoiding them. The various results expected were to form a

basis for a future operational capability of our system.

1.2 Method

Linguists and other humans are notoriously ill-equipped to recognize sentential ambiguities. As an example we may cite the by now hopefully infamous example *time flies like an arrow*. It took approximately 3 years to increase the number of recognized ambiguities from three to four. Since the emphasis of our interest was on the resolution of sentential ambiguities, we needed to make sure that such ambiguities were recognized. The method we planned to follow was thus basically descriptive. Computer analyses were to be provided of a given text. The sentences with ambiguous interpretations or multiple translations were to be classified as to the type of ambiguities, lexical or structural. We were then going to determine why, as human beings, we were able to select the proper reading of the sentence, that is, disambiguate it. The information we based our decision on was to be divided into several sections:

a. information which was contained in the sentence but was not evaluated or used by the algorithm;

b. information contained in the immediately preceding and/or following textual environment of the sentence in question;

c. information contained in the text other than in the immediate environment, and

d. other information.

Obviously, information under a. was to contribute to the continuing development of our lexicographic classification. The disambiguating information was to be determined by simulation of the performance of the connection component described in "Requirements...", that is, by determining the various logical relations holding between the presuppositions, assertions, and entailments of the lexical pieces and sentences in question.

1.3 The Corpus

The test corpus for our experiment, the monograph on rocket propulsion systems, *Raketenantriebe*, by Joseph Stemmer [8], consisted of 523 pages of text and figures. The text contained 135,266 running words; the number of distinct words was 20,466. We selected the first 3 chapters (approx. 150 pages) of the text for our analysis. The number of running words in these chapters were 45,134, the number of distinct words, 9,111. The first chapter, a historical review, consisted mainly of sentences containing dates, names, and only few predicates, which made it ideal for testing the system's capability to analyze and translate

non-well-formed text. The second and third chapters contained regular prose.

1.3.1 Lexicographic Description of RAK

The lexicographic work, the establishment of English translation equivalents, and the classification of German and English dictionary items, was performed for the whole corpus. The details are described in Section III.

1.3.2 Syntactic Description

In order to account for every syntactic structure occurring in Sections II and III, a narrow description of the clause level structure of each sentence was provided. These descriptions were subsequently punched and processed by means of a glossary program. Section II consisted of 1,016 sentences with 1,722 clauses, 160 of which contained deleted elements. The number of distinct clause patterns was 918. Section III contained 1,090 clause patterns, 682 of which were distinct.

These lists were processed by means of conversion programs which generated a broader description and reduced iterations of symbols to one symbol (thereby simulating to some extent occurrences of optional symbols). The 984 distinct patterns of Section II could thus be reduced to 534 distinct patterns. By means of a subsequent program those patterns were alphabetized, thereby simulating rules of a permutation grammar. This reduced the number of distinct patterns to 159.

It became evident, after some time, that the modification of the syntactic analysis algorithms to process optional symbols or even set-theoretic rules would not be realized before the actual demonstration of the system's capability (cf. Section 1.4, below). We thus reduced the amount of text to be processed to what we hoped was still a representative sample— three pages with 42 sentences, consisting of 72 main and subordinate clauses interpreted by 86 distinct clause rules. (The greater number of clause rules reflects the multiple descriptions assigned to ambiguous sentences.)

1.4 Execution of Experiment

During this contract period, the first part of the experiment was successfully completed. The main components of LRS were constructed. The grammar formalism was modified to incorporate the capability for the generation of multiple translations for output sentences (cf. Appendix I). The various grammatical descriptions and lexical items necessary for the translation of the text were provided.

The second part of the experiment, a detailed evaluation of possibilities and limitations of sentence-by-sentence translation will be given in a future report.

As we recall, the objective of the experiment was to show the extent to which an MT system performing sentence-by-sentence translation would produce high quality output. Because the format of the standard output as originally proposed might not have been visually acceptable to the non-specialist user (cf. Fig. I.1), additional components had to be developed to improve the legibility of the output; hence, the output syntactic choice component and the lexical spellout component. They produce a version of English standard text which we call "user standard".

Actual Text

Soon after the physiologists had demonstrated that conduction in the sensory nerves takes place by all-or-none firings— a king of impulse code— the question arose whether, under the proper circumstances, we can perceive an all-or-none increase or decrease in sensation. In other words, does a smooth increase in a stimulus produce finite jumps in perception? No such step functions had yet been reported in 1926 when Boring hypothesized what he called Sensory Quanta and speculated about their relation to the sensory continuum. "It should be perfectly feasible experimentally", he wrote, "to determine the presence or absence of critical points..." (1)
From *Science*, Vol. 177, Number 4051, 1 Sept. 1972.

Simulated standard English of text above:

SOON AFTER THE PHYSIOLOGIST-S HAD PPL DEMONSTRATE THAT CONDUCTION IN THE SENSORY NERVE-S DOES TAKE PLACE BY ALL-OR-NONE FIRING-S (A KIND OF IMPULSE CODE) THE QUESTION DID ARISE WHETHER, UNDER THE PROPER CIRCUMSTANCE-S, WE CAN PERCEIVE AN ALL-OR-NONE INCREASE OR DECREASE IN SENSATION. IN OTHER WORD-S, DOES A SMOOTH INCREASE IN A STIMULUS PRODUCE FINITE JUMP-S IN PERCEPTION? NO SUCH STEP FUNCTION-S HAD YET BEEN PPL REPORT IN 1926 WHEN BORING DID HYPOTHESIZE WHAT HE DID CALL SENSORY QUANTA AND DID SPECULATE ABOUT THEIR RELATION TO THE SENSORY CONTINUUM. "IT SHOULD BE PERFECTLY FEASIBLE EXPERIMENTALLY," HE DID WRITE, "TO DETERMINE THE PRESENCE OR ABSENCE OF CRITICAL POINT-S..." (1)

Fig. I.1

The actual work on the experiment, the construction of the systems components and grammars, was initially considerably delayed due to the fact that only insufficient funds were provided for computer time. That we were able to perform the demonstration of the systems capability June 1974, we owe to a generous contribution of computer time to the Computation Center of the University

of Texas. This contribution also permitted us to submit our analysis jobs in individual portions rather than as complete runs. The increase in cost based on the requirement of saving intermediate output was outweighed by the increase in turn-around which expedited debugging of programs and grammars. That the turn-around time was generally still slow was primarily due to the system's components requiring a core storage of 220K, the maximum allowed to each user. With developments in the SVOL package— the central routine of every analysis program— some of the programs exceeded this limit and could only be run by a special arrangement with the Computation Center on Sunday mornings. Certain of the system's components eventually required a core storage of 260K words, which impeded other normal operations of the Computation Center.

1.5 Results

An intermediate evaluation of the performance of LRS shows the following results. The output of sentence-by-sentence translation by our system is surprisingly good (cf. Appendix I). The sentences were disambiguated to a large extent solely with the information contained in that sentence. (In addition, the number of multiple translations produced by the system can be further reduced by omitting from the output language those translation equivalents felt to be synonymous with some other term.) The system recognizes potential sentence boundaries satisfactorily, and thus permits translation as a fully automatic process requiring no pre- or post-editing.

In view of the small size of the translated text, no reliable answers can yet be given to many of the questions which had initiated this experiment. The size is, however, sufficient to easily confirm that in cases where the necessary disambiguating semantic information is not stated in a particular sentence, a mechanical translation system based on sentence-by-sentence translation cannot solve the problems of establishing anaphoric relations or pronominal reference, or disambiguating lexical or structural ambiguities. The experiment also showed convincingly that the speed of the processing algorithms, and thereby the cost factor of mechanical translation, is dependent on the volume of the linguistic information that needs to be handled during the analysis and translation process. Moreover, the speed of producing mechanical translation is dependent on the number of components which have to be used during the actual translation process.

We plan to continue our research in the areas of the three criteria used for the evaluation of MT, quality, cost and speed. We plan to improve the quality even further by designing a contextual component to permit a solution to problems posed by deletion recovery, anaphoric relationships, and ambiguity solution. We need to decrease the cost of operation by compacting the grammatical information used by the programs (lexicon), by reducing the number of rules necessary for analysis (permutation grammar

with optional symbols), by permitting an ordered application of the rules during the analysis and production process, and by making rule application sensitive to the analyses obtained so far.

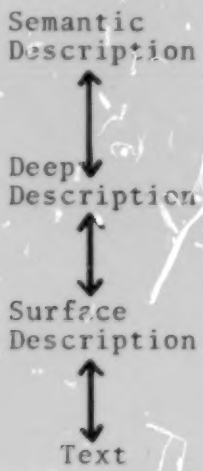
We plan to decrease the number of processing stages by making some components optional like standardizing component or the lexical collocation component. They will only be called if the text actually requires it, as, for example, in the case of lexical collocations. A further reduction in processing stages can be obtained by increasing the power of some components which will make others unnecessary. Thus the output choice mechanism will be improved to permit the generation of English surface strings directly from the English standard representation rather than going through a separate rearrangement phase. The effect of these changes on quality, cost and speed will be reported on in future reports.

SECTION II

2. The Linguistics Research System (LRS)

2.0 Linguistic Basis

LRS is a configuration of algorithms and grammars designed to derive from a given sentence P the meaning(s) of P and given a particular meaning m to derive all sentences with that meaning. Like other grammatical models, it assumes three levels of representations of a sentence: the surface description, the deep description, and the semantic description. We call the latter two "standard readings" and "normal form readings", respectively. The following graph represents these levels of representations.



Note that the arrows are bi-directional; that is, LRS is a system of effective procedures for the derivation of the meaning of a text from the text, and the derivation of a text from the meaning of that text.

LRS assumes that the meaning of a sentence is independent of the language in which the sentence is uttered; that is, LRS assumes a universal semantic base. Every arrow in the graph above corresponds to a pair of grammars and algorithms which process these grammars. The up-arrow corresponds to the input language pair, the down-arrow, to the output language pair. If the grammars used for input and output language are identical, we call the process paraphrasing; if the grammars are distinct, we call the process translation.

Linguists have argued convincingly in the past that mechanical translation from surface structure directly to surface structure cannot produce satisfactory translation, and that a deeper repre-

sentation has to be derived first which can then be used to translate into a corresponding deep representation of the output language. To validate this argument it may be sufficient to point out that often translation requires a change in part of speech of the translation equivalents. For example, German

er trinkt gern

where the German adverb *gern* becomes the English verb *like*

he likes to drink; or

die Bundesrepublik Deutschland

the Federal Republic of Germany

where the German noun *Bund* is translated as the English adjective *Federal.*

Similarly, the occurrence of lexical collocations requires a translation performed on a level deeper than the surface description. Lexical collocations are dictionary entries which consist of more than one word and whose syntactic interpretation and meaning cannot be derived from the syntactic interpretation and meaning of the individual words of the collocation (example: *in Frage stellen = jeopardize*, or in English: *make up one's mind = decide*). The frequency of their occurrence is higher than generally assumed. We discovered 135 distinct lexical collocations in the about 1,500 sentences of Chapters II and III of our test corpus; also cf. Lexicography, below.

2.1 General Description of LRS Grammars

Whenever we use the expression "grammatical description" in LRS, we are really talking about two grammatical descriptions simultaneously. The first description is the one provided by the linguist. These descriptions are stated in very general terms, for example, for German:

> a noun phrase may consist of a determiner followed by a noun; the two constituents must agree in gender, number and case. The noun phrase acquires this common gender, number and case; moreover, the noun phrase has all the properties of the noun.

The grammar formalism permits the statement of such general syntactic and semantic descriptions by means of the set-theoretical operations intersection, union, and complement. These are used to express relations which may hold between the values of specified subscripts.

The grammars written by the linguists are context-free grammars with complex symbols.

The corresponding processing algorithm operates with the grammatical description given by the linguist and generates the actual grammatical description of a constituent. The description provided by the algorithm is far more precise because it mentions all the properties of the constituents which are not explicitly mentioned in the rule.

Since one grammar rule thus represents an abbreviation of a large but finite number of descriptions provided by the algorithm, the advantages of using such grammars with complex symbols are obvious. Certain disadvantages of context-free analysis are, however, not overcome by complex-free grammars. Thus the constituents to be interpreted by a rule have to occur in a text in the order specified by the rule. This results in a proliferation of rules for languages which permit limited or free permutation of their constituents. (A comprehensive description of the grammar formalism and their linguistic descriptions is given in Appendices II and III.)

2.2 General Description of LRS Algorithms and their Interrelationship with the LRS Grammars

With two exceptions the LRS algorithms are direct substitution analyzers. They perform grammatical analysis from bottom-to-top, processing a text from left-to-right. At each text position T they apply all rules which provide an analysis for every text span beginning to the left of T and including T. Such analyzers have been described in detail at various places. In this description of the LRS algorithms we will mainly point out the function of each algorithm, the grammar it uses, and the interrelationship between grammatical description and algorithmic performance.

The subsequent passages will, wherever applicable, show the following structure. The function of the algorithm is described, followed by the reasons for the algorithm's performance, if necessary. We then briefly describe the grammatical instructions upon which the algorithm operates. We then discuss the mutual effects between algorithms and grammars. Each subsection ends with a graph showing the results of the analysis by that component of a German sentence which occurred in our text:

Damit tritt das fundamentale physikalische Gesetz

<<Actio et Reactio>> in Kraft.

Literally:

Thereby kicks the fundamental physical law <<actio

et reactio>> in/into force.

Standard English:

The fundamental physical law <<actio et reactio>> thereby becomes effective.

2.3 The Surface Component

Surface Analysis accepts a text of the input language and processes it with the grammars for that language to produce as output a set of standard sub-trees—in former publications referred to as tentative standard strings.

Surface Analysis is executed in three phases, the dictionary phase, the word phase, and the syntactic phase. Each of these phases consists of an analysis stage and a choice stage (cf. Fig. II.1). The surface component thus consists of six sub-components: the dictionary analysis component, the dictionary choice component, the word analysis component, the word choice component, the syntactic analysis component, and the syntactic choice component.

[For Fig. II.1 refer to following page]

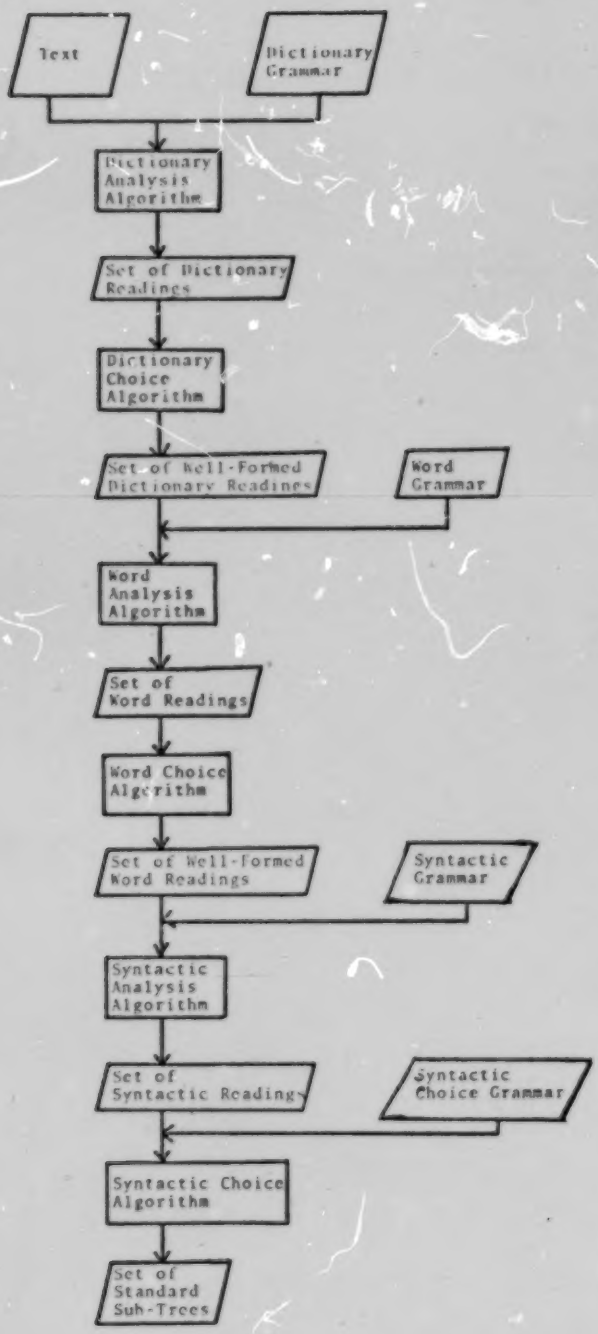

Text
Dictionary Grammar
Dictionary Analysis Algorithm
Set of Dictionary Readings
Dictionary Choice Algorithm
Set of Well-Formed Dictionary Readings
Word Grammar
Word Analysis Algorithm
Set of Word Readings
Word Choice Algorithm
Set of Well-Formed Word Readings
Syntactic Grammar
Syntactic Analysis Algorithm
Set of Syntactic Readings
Syntactic Choice Grammar
Syntactic Choice Algorithm
Set of Standard Sub-Trees

Figure II.1

2.3.1 The Dictionary Component

2.3.1.1 Dictionary Analysis

The purpose of the Dictionary Analysis Component is to find all sequences of contiguous stems and endings which cover a word, beginning with the first letter. In the word

GROESSEREN

the stems and endings underlined will thus be recognized.

The reason for recognizing all stems and endings is to account for the ability of a number of languages to form compounds. German is well known, if not infamous, for its noun compound formation which theoretically permits the generation of an infinite number of nouns. This particular property prevents the application of the longest-span method used for languages like English, French, and Russian at such institutes as TAUM (Montreal) and GETA (Grenoble). Such a procedure would lead to different results for German, depending on whether dictionary analysis was performed from left-to-right or right-to-left. The word *VORKRIEGSPIONIEREN* would thus either be segmented into *VORKRIEGS + PIONIEREN* (left-right), or *VORKRIEG + SPIONIEREN* (right-left).

Not all stems or endings which occur in a word are recognized by the algorithm. Thus the sequence *ESSER* or the letter *E* (after *GRO*) are not recognized as a noun or an ending respectively, since no stem or ending terminated at the letter *O* in *GROESSER*. The recognition of such intermediate spans as dictionary entries is prevented by means of preconditions and post-conditions which can be associated with each dictionary entry. Preconditions are of four kinds:

a. the predecessor must have been a marginal symbol (blank and/or punctuation mark);

b. the predecessor must have "set a morpheme boundary";

c. the predecessor must have been of type a or type b;

d. the predecessor is ignored.

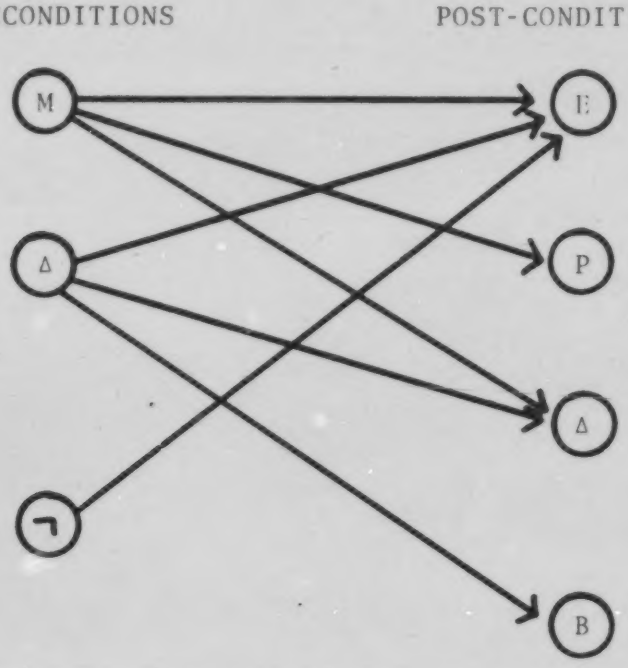
Post-conditions associated with the dictionary entry have to be satisfied by the preconditions of the rules applying to the string following such an entry. The restrictions which can be set are:

e. set a marginal symbol boundary;

f. set a morpheme boundary (stem or ending boundary);

g. do not set a morpheme boundary.

The transition possibilities between preconditions and post-conditions are shown in the following graph.



The arrows indicate: a rule with precondition X can be followed by a rule with post-condition Y if an arrow leads from X to Y.

M = identifies "marginal symbol" (blank space, punctuation marks)

Δ = as precondition: identifies a morpheme,
as post-condition: must be preceded by marginal symbol or morpheme

¬ = identifies an element which can only be followed by a marginal symbol

E = identifies an element which can occur anywhere (attached to blank space and punctuation marks)

P = identifies an element which must be preceded by marginal symbol

B = identifies an element which must be preceded by a morpheme (attached to suffixes).

Thus, the letter G does not set a morpheme boundary which prevents *ROESSER* (*horses*) from being recognized, since *ROESSER* requires a precondition of type b. The same condition prevents *ESSER* from being recognized. The letter *E* (after *O*) has a precondition of type b; the pronoun *ER* requires a precondition of type a.

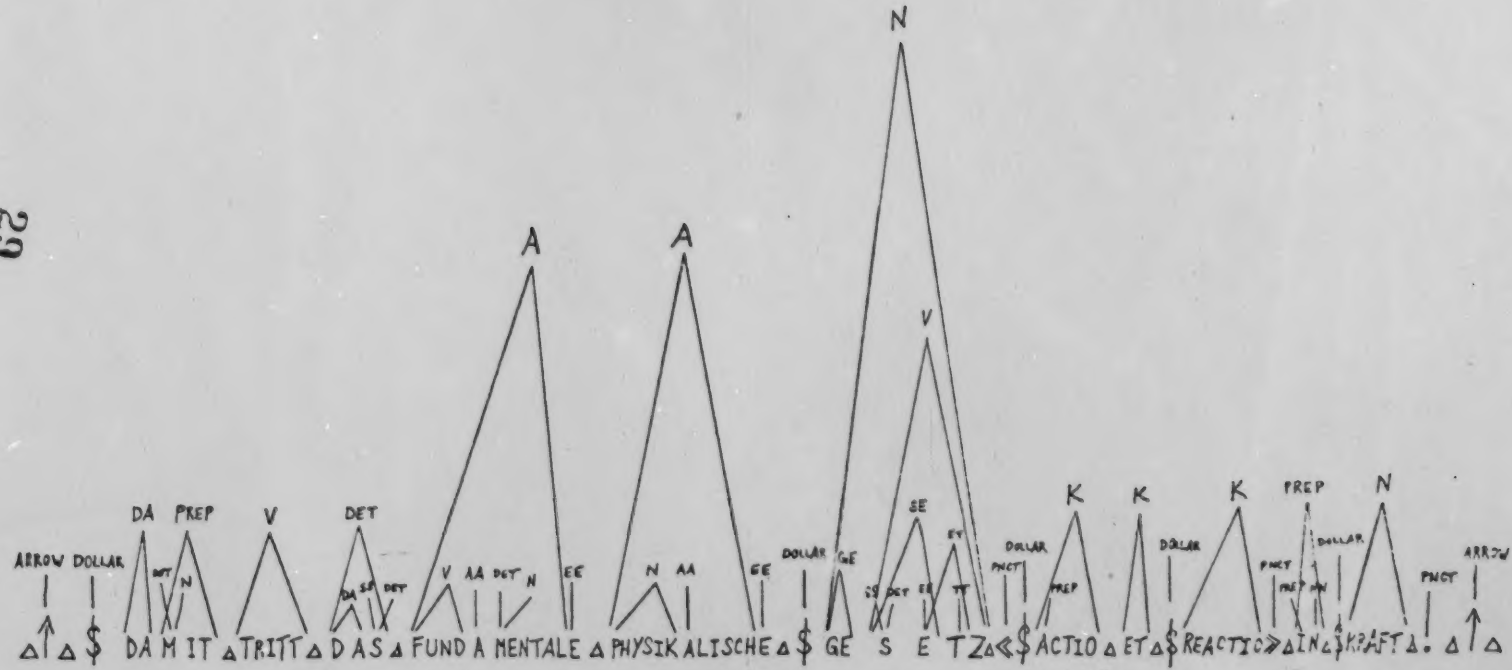
Dictionary Analysis is performed with the dictionary grammar which is stored as a set of 64 trees, the roots of which represent the initial symbol a dictionary entry can begin with. Thus all words beginning with A occur in the tree with root A, those beginning with B occur in the tree with root B, etc. Each root is connected through n branches with daughter nodes where each such daughter node represents a letter which occurs as the second letter after the "root" letter in a dictionary entry. An additional daughter node is added after each node which ends a word.

The analyzer follows and discovers all possible interpretations "simultaneously". Thus, when the analyzer discovered the adjective stem *GROESS* in the example above, it continued the analysis which eventually discovered the noun *GROESSE* but at the same time started the analysis which discovered the endings *E* and *ER*.

[For Fig. II.2 refer to following page]

Fig. II.2

S 16
DICTIONARY ANALYSIS


ARROW DOLLAR DA PREP V DET A A N V SE K K K PREP N
DET N DA SE DET V AA DET N EE N AA GE DOLLAR GE SE DET EE TT PNCT DOLLAR PREP DOLLAR PNCT PREP DOLLAR PNCT ARROW
Δ↑Δ$ DA M IT ΔTRITT Δ D AS Δ FUND A MENTALE Δ PHYSIK ALISCHE Δ$ GE S E TZΔ«$ ACTIO Δ ET Δ$ REACTIO»Δ IN Δ$ KRAFT Δ. Δ↑Δ

2.3.1. 2 Dictionary Choice

The purpose of Dictionary Choice is to discover all stems and/or endings which cover a word contiguously, beginning with the first letter of a word and ending with the last letter of a word. Entries which are not part of such a contiguous complete sequence are deleted. Our representation of dictionary analysis of

GROESSEREN

results after Dictionary Choice in

GROESSEREN.

Besides computing the dictionary entries to be retained for subsequent analysis, Dictionary Choice can be influenced by a preference operator, given in certain dictionary entries. The effect of a preference operator is that all well-formed subsequences within a dictionary entry are deleted. Assume that the dictionary contains *RUECKSTOSS*, *STOSS*, *KRAFT* and *STOSSKRAFT* where *STOSSKRAFT* has a P operator. If the text had *STOSSKRAFT*, the entries *STOSS* and *KRAFT* would be deleted after Dictionary Choice. If *RUECKSTOSSKRAFT* occurs in a text, however, the entries for *RUECK*, *STOSS*, and *KRAFT* remain after Dictionary Choice.

This Preference Operator was mainly attached to rules for German endings in order to prevent the occurrence of intermediate, parasitic interpretations. Thus the Preference Operator attached to the ending *EN* deleted the endings *E* and *N* in *GROESSEREN*, above, preventing an interpretation of *GROESSERE* during word analysis.

Spans which are not covered by a complete sequence of interpretations are assigned a K-reading, which stands for "unknown word(s)" (cf. *actio*, *et*, and *reactio*, Fig. II.3, below).

[For Fig. II.3 refer to following page]

Fig. II.3

S 16
DICTIONARY CHOICE
ARROW
DOLLAR
DA
PREP
V
DET
A
EE
A
EE
DOLLAR
GE
N
V
PNCT
K
K
K
PNCT
PREP
DOLLAR
N
PNCT
ARROW
Δ↑Δ$DAMITΔTRITTΔDASΔFUNDAMENTALEΔPHYSIKALISCHEΔ$GE SETZΔ<$ACTIOΔETΔ$REACTIO»Δ IN Δ $ KRAFTΔ.Δ↑Δ

2.3.2 The Word Component

The purpose of the Word Component is threefold:

a. to find all segments or sequences of segments which are well-formed according to the word grammar of a language, and record that description;

b. to delete all segments or sequences of segments which are not well-formed;

c. to determine those text positions where a sentence or clause boundary could theoretically occur;

d. to identify the nodes on which syntactic analysis is to continue building.

The reasons for (a) are obvious.

The incorrect segments are removed (b) to reduce processing time and to increase the amount of work space and grammar information that can be kept in store during processing.

Tentative sentence boundaries (c) were introduced for the following reason: a direct substitution analyzer without any restrictions will provide two sentence readings for the text *the experiments succeeded*: one for *experiments succeeded* and another for *the experiments succeeded*. By making clause rules and sentence rules sensitive to potential sentence boundaries, the clause and sentence interpretations could be restricted. The sequence *the demonstrations and the experiments were performed* would thus show only two analyses as sentences, namely for the span *the experiments were performed* and for the whole text; these two analyses cannot be reduced to one since the word *and* as well as the "." were used to establish potential sentence boundaries.

The restriction of constituents which could be built on by syntactic analysis (d) was obtained by flagging all nodes except for the top node of a constituent, thereby making them "invisible" to the syntactic analyzer. A sequence like *had been seen* would thus be interpreted as a predicate and none of the intermediate structures would be available for subsequent syntactic analysis.

2.3.2.1 Word Analysis

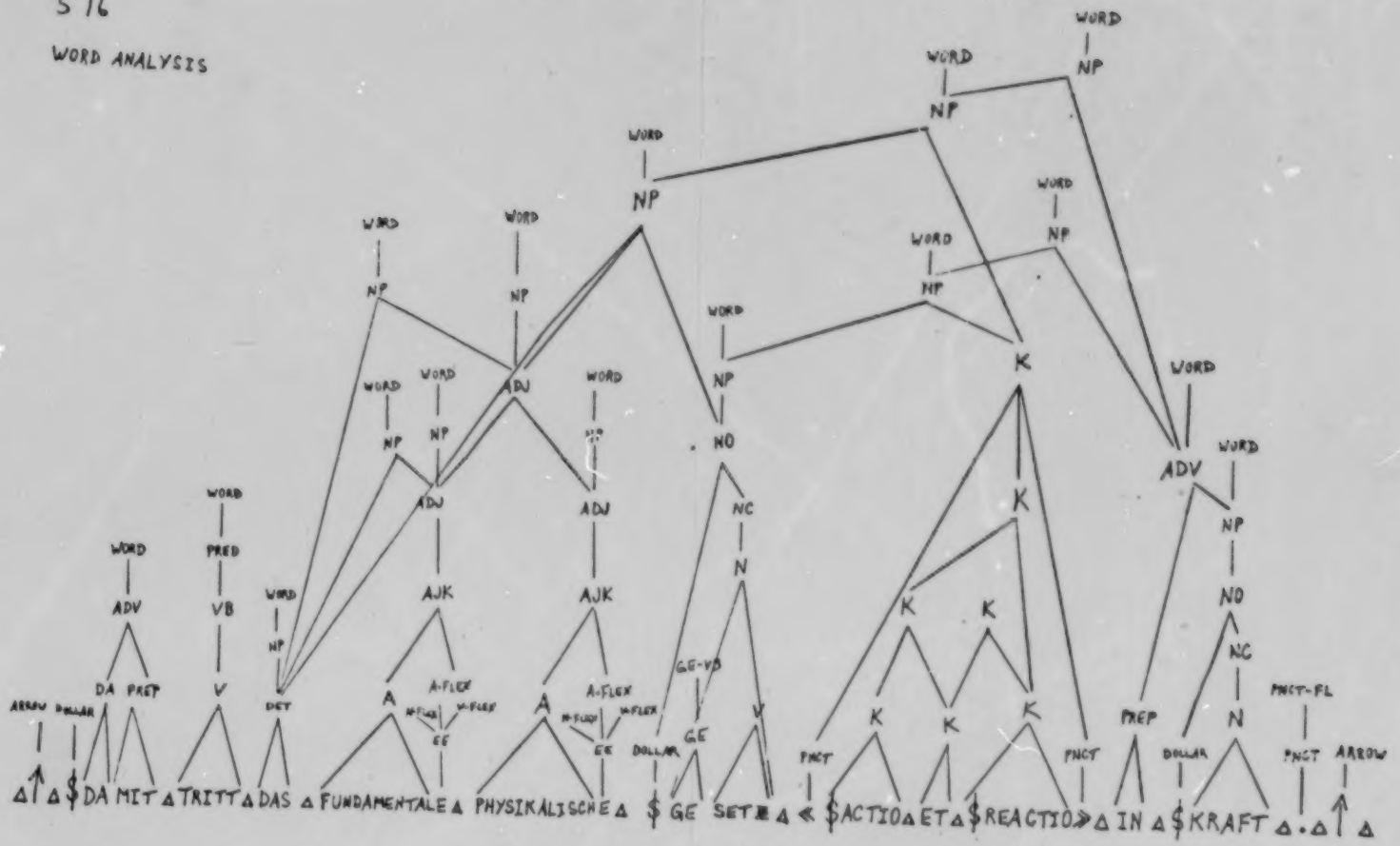
Word Analysis applies all rules applicable to every text span available at each text position in the text. The application of these rules can be restricted by means of the operators P and F which can be attached to constituents in a rule consequent. The operator P indicates that the constituent must be preceded by a marginal symbol, the operator F stands for

"constituent must be followed by a marginal symbol". By means of these operators, a large number of parasitic interpretations occurring during the morphological interpretation of a word can be avoided. Thus, the span *GROESSE* occurring in the word *GROESSEN* is not interpreted as a nominal (a noun + its inflection endings) since it is not followed by a marginal symbol. Word Analysis operates with the word grammar of a language, whose initial symbol for all languages is V WORD. Any text span covered by an analysis dominated by V WORD is for that reason well-formed according to the word grammar of that language.

The speed with which Word Analysis (and also Syntactic Analysis) were executed had a considerable effect on the grammar design which will be described below under Syntactic Analysis.

[For Fig. II.4 refer to following page]

Fig. II.4


S 16
WORD ANALYSIS
ARROW DOLLAR DA PRET VB DET A A-FLEX M-FLEX E-E GE-VB GE V K ADV PREP NC N PNCT-FL
Δ↑ Δ $DA MIT Δ TRITT Δ DAS Δ FUNDAMENTALE Δ PHYSIKALISCHE Δ $GE SETZ Δ « $ACTIO Δ ET Δ $REACTIO » Δ IN Δ $KRAFT Δ . Δ ↑ Δ

2.3.2.2 Word Choice

Word Choice performs the functions under (a) through (d) above

a. by retaining all readings dominated by WORD;

b. by destroying all interpretations not dominated by WORD, including the artificial symbol WORD;

c. by inserting potential sentence boundaries into the text spans according to the specifications of the word analysis rules applying to that text span; and

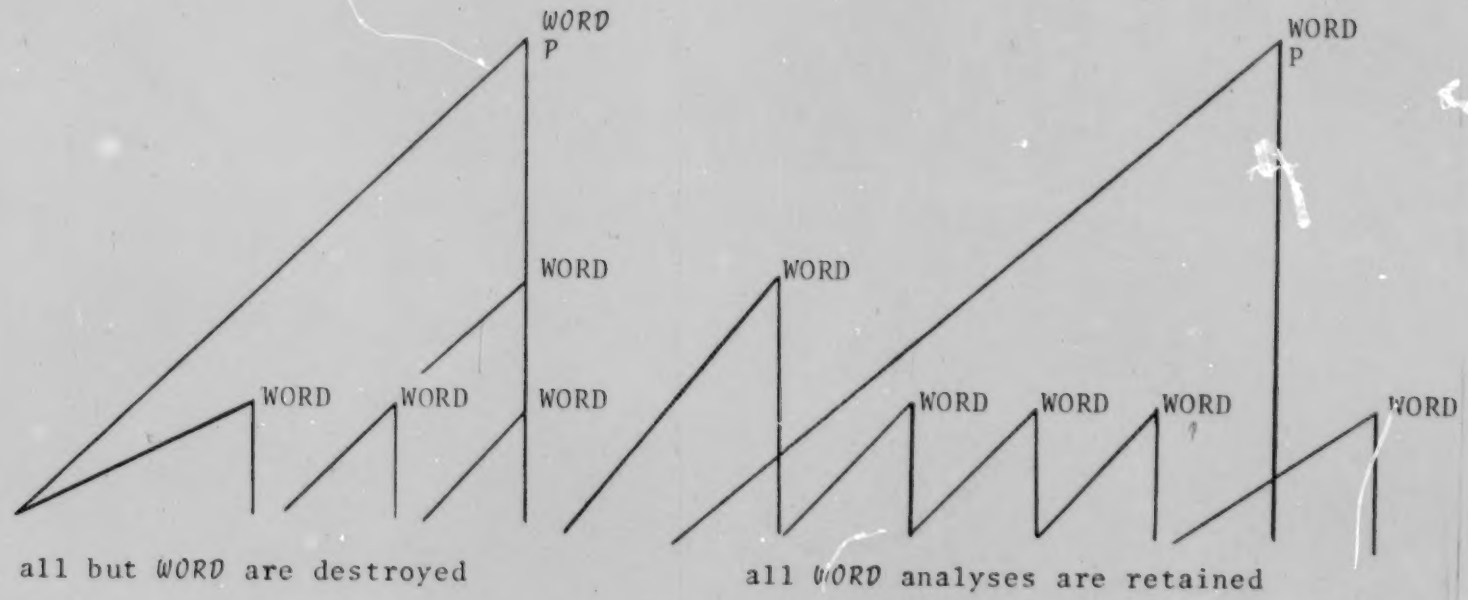
d. by flagging every node which was not immediately dominated by WORD.

In order to permit Word Choice to execute these functions more efficiently, special operators can be attached to word rules. The D-operator permits the destruction of a node and of every node building on it. It was used to destroy interpretations covering a word which contained internal syntactic boundaries like the German *am* which consists of a preposition followed by a determiner.

The S-operator permits the attachment of "super flags" to nodes directly dominated by V WORD. The I-operator permits the insertion of new subscripts and values into a constituent. It was thus possible to attach information pertaining to the context in which a word occurred to the constituent, and thus restrict its re-writing during the subsequent syntactic analysis phase.

The most powerful operator is the P-operator (Preference Operator). It can only be attached to the constituent WORD. It permits the algorithm to destroy every constituent in a text span not dominated by this particular symbol WORD. The P-operator is executed based on the longest span principle: all readings WORD applying to text spans contained in a P-span are deleted even if they have a P-operator of their own. Text spans which are only part of a P-span are retained as well as all spans contiguous to the retained spans (cf. Figures II.5 and II.6).

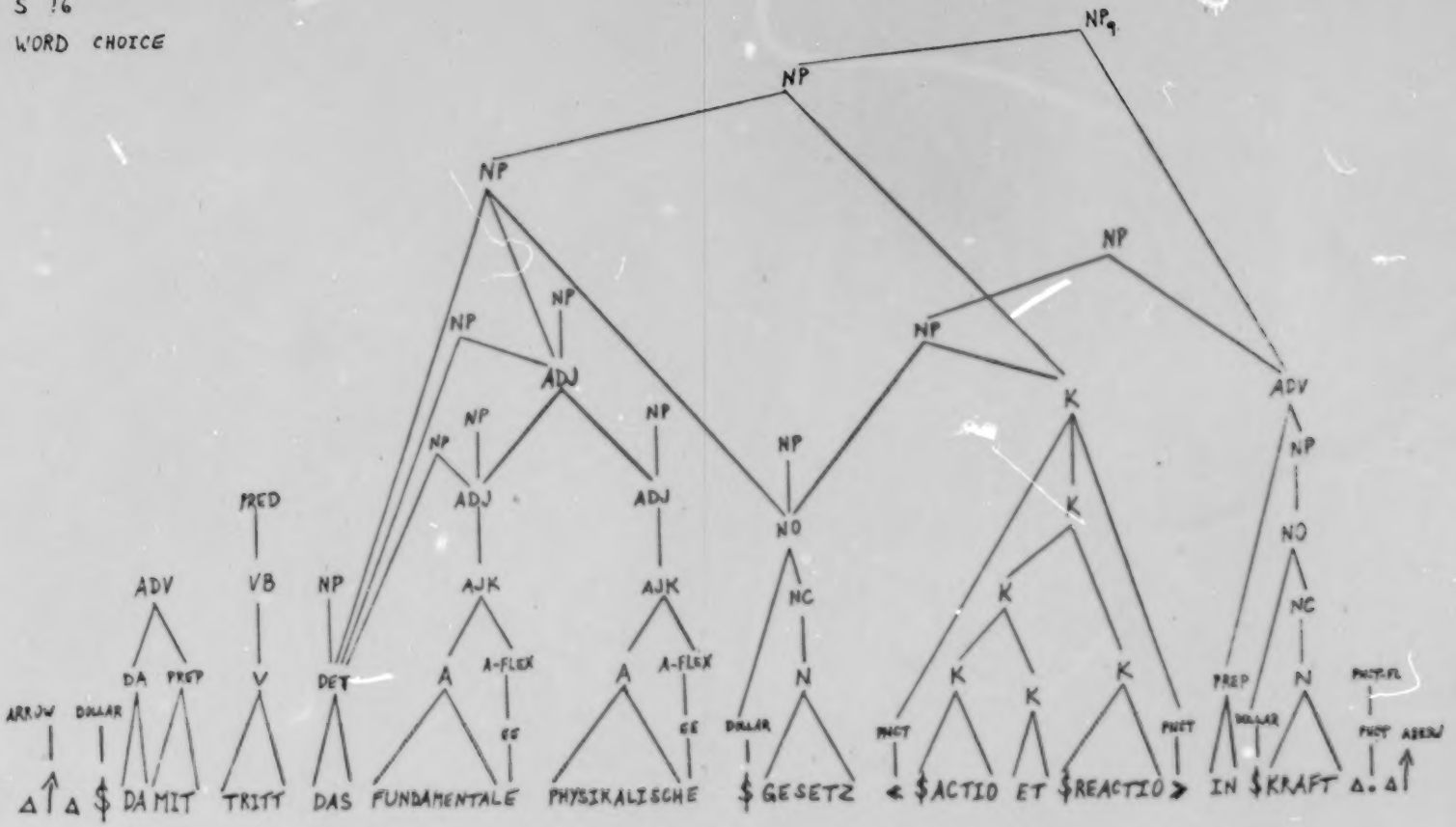
[For Fig. II.5 refer to following page]
[For Fig. II.6 refer to following page]



V-WORD spans destroyed and not destroyed within P-spans.

Fig. II.5

Fig. II.6


S 16
WORD CHOICE
NP
ADJ
ADV
VB
PRED
DET
AJK
A
A-FLEX
N
NC
K
PREP
DA
V
DOLLAR
ARROW
PUNCT
ABBR
DA MIT
TRITT
DAS
FUNDAMENTALE
PHYSIKALISCHE
$GESETZ
<$ACTIO ET $REACTIO>
IN $KRAFT
Δ. Δ

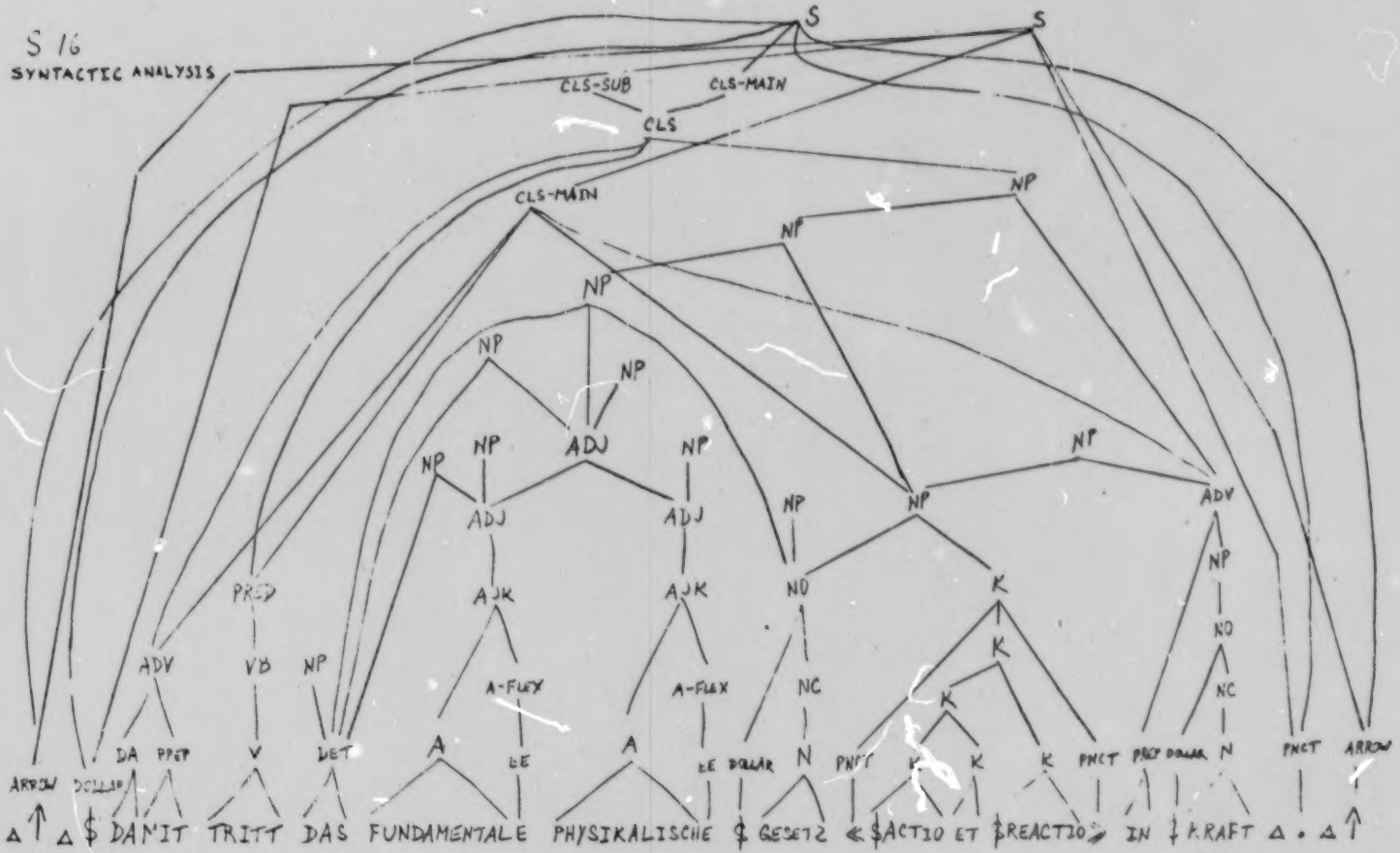
2.3.3 The Syntactic Analysis Component

2.3.3.1 Syntactic Analysis

The purpose of Syntactic Analysis is to find all readings of a text span which is well-formed according to the syntactic grammar of a language. These readings are dominated by the grammar's initial symbol S. The Syntactic Analysis algorithm is now identical to the Word Analysis algorithm. It operates in one of two modes: in the first mode it builds on any reading, whether flagged or unflagged (like Word Analysis); in the second mode it only builds on unflagged readings (the original Syntactic Analysis algorithm). The Syntactic Analysis algorithm operates with the syntactic grammar of a language. This grammar changed considerably during the period the experiment was in preparation. A detailed description of its current version is given in Appendix II. In the following section we will only discuss basic changes necessitated by the performance of the (Word and) Syntactic Analysis algorithm.

[For Fig. II.7 refer to the following page]

Fig. II.7


S 16
SYNTACTIC ANALYSIS
S
S
CLS-SUB
CLS-MAIN
CLS
CLS-MAIN
NP
NP
NP
NP
NP
NP
NP
NP
NP
NP
NP
ADJ
ADJ
ADJ
ADJK
ADJK
ND
K
K
K
K
K
ADV
NP
ND
NC
PRED
VB
NP
A-FLEX
A-FLEX
NC
ADV
VB
NP
DET
A
EE
A
EE
DOLLAR
N
PNCT
K
K
K
PNCT
PREP
DOLLAR
N
PNCT
ARROW
ARROW
DOLLAR
DA
PREP
V
DET
Δ ↑ Δ $ DAMIT TRITT DAS FUNDAMENTALE PHYSIKALISCHE $ GESETZ « $ACTIO ET $REACTIO IN $KRAFT Δ . Δ ↑

2.3.3.2 Restrictions and Changes - Surface Grammar

We mentioned in Section I that the necessity of meeting deadlines prevented the completion of certain parts of the Surface Word and Syntactic Analysis algorithms. This had a pronounced effect on the amount of data (grammar rules) that had to be processed by them and consequently on their speed. Three factors were primarily involved.

1. The algorithm did not conflate identical constituents covering the same span. This resulted in a proliferation of a number of clause and sentence interpretations. For example, consider the relative clause *who saw the boy with a telescope*. This has two interpretations, one with the prepositional phrase is analyzed as a post-nominal attribute, and one with it as a verb modifier. The phrase *the man who saw the boy with a telescope* would thus be analyzed twice. The first time, concatenating *man* and the first interpretation of the relative clause; the second time, *the man* with the second interpretation of the relative clause. The sentence *he talked to the man who saw the boy with a telescope* would consequently also be analyzed twice as a clause and receive two S readings, rather than once as a clause with a noun phrase which dominated an inherently ambiguous relative clause. Since subordinate clauses are interpreted as adverbials of their matrix clauses, obtaining a large number of such additional analyses was often unavoidable. (The maximum number of analyses obtained was 24 S readings for a sentence.) Though the Syntactic Choice grammar (cf. below) permits the linguist to select between such readings on the basis of linguistic and/or pragmatic considerations, we found it necessary— in order to improve the analysis speed— to use ad hoc restrictions in the corresponding grammar rules in order to restrict the number of prepositional phrases which could function as post-nominal attributes. In effect, only the German prepositions *vom, im, mit* and *zwischen* were allowed to occur as heads of post-nominal prepositional phrases.

2. We already mentioned the algorithm's inability to process set-theoretical rules or rules with optional symbols. For these reasons the size of the text analyzed for the experiment had to be considerably reduced because the number of clause rules necessary for the syntactic description of the original three chapters became unmanageably large (cf. Section I).

3. Before the algorithm applies a rule it performs basically three checks. It first determines whether the category symbols of a sequence of nodes correspond to the category symbols in the consequent of a rule. It then checks whether each node has the subscripts and values specified in the rule. Third, it performs the operations specified in the rule. Whenever these three checks are successful, it constructs the new constituent corresponding to the instructions in the antecedent of the rule.

When the algorithm compares text category symbols against

rule category symbols, it constructs intermediate tables in which it records the results of the matching process obtained so far. Because of the generality of the category symbols in the Word and Syntactic grammars, the number of attempted rule applications and the size of the intermediate tables became so large that the maximum time allowed the algorithm for the construction of a rule had to be increased several times. In order to reduce the size of the intermediate tables we were forced to make the category symbols more specific which, in some cases, caused a duplication of rules. For instance, there was the general rule which conjoined, as a single noun phrase, the sequence of a noun phrase followed by a conjunction followed by a noun phrase (NP→NP CONJ NP). This had to be separated into 1) a noun phrase followed by comma followed by noun phrase, and 2) a noun phrase followed by conjunction other than comma followed by noun phrase. The rules most affected by this decision were the morphological rules, where adjective endings, noun endings and verb endings now had to be distinguished and the clause rules (for more details, cf. Appendix II).

Fortunately, making the category symbols in clause rules more precise did not increase the number of clause rules, since this change only affected the rule antecedent. We established four types of clauses: V CLS-MAIN, V CLS-SUB, V CLS-HYP, and V CLS (for main clause, subordinate clause, hypothetical or interrogative clause, and "undetermined" clause, respectively). The latter was eventually rewritten as one of the first three clauses, depending on the subscripts and values associated with it. V CLS-SUB was rewritten as V CLS-REL (relative clause) if its initial constituent dominated a relative pronoun.

The former rule (1)

```
V NP      →   V NP        V CONJ          V CLS
∧ 2               2       $ TY(COMMA)     $ TY(REL)
```

was thus changed to (2)

```
V NP      →   V NP        V COMMA         V CLS-REL
∧ 2               2
```

With the first rule, the algorithm only determined during the subscript and value check that the sequence *the girl, he smiled* contained in *When he saw the girl, he smiled*, was not a sequence of a noun phrase followed by a conjunction followed by a relative clause. With the increased amount of information contained in the category symbols of rule 2, the rule is not even tried. These changes led to a considerable reduction in analysis time. They did not, however, alleviate another problem of the analyzer. During the analysis of long sentences (consisting of more than 200 letters) which contained a relative clause, the number of the intermediate tables became so large that the time allotted for the construction of a new constituent was frequently exceeded. We were therefore forced to sub-divide the syntactic grammar into

two parts:

1. phrase grammar

and

2. clause grammar

Each grammar was processed during a separate analysis phase. It turned out, however, that the addition of another analysis component with its loading of programs, loading of grammars, etc. required more time than could be saved by sub-dividing the grammar. In the final solution, we extended the scope of the word grammar to include phrasal constructions such as noun phrases, prepositional phrases, their conjuncts and expansions by post-nominal modifiers. The effect was that the size of the intermediate tables was roughly divided in half, expediting the combination of Word and Syntactic Analyses.

More than 90% of the current syntactic grammar consists of clause rules, rules which concatenate noun phrases and relative clauses, and rules which rewrite a relative pronoun as a noun phrase or prepositional phrase and sequences of one-word adverbs as adverbials.

2.3.3.3 Syntactic Choice

The primary purpose of Syntactic Choice is to disambiguate words on the basis of information contained in a sentence, and to determine the semantic function of each constituent and its semantic relations towards other constituents. In the process of making these decisions, Syntactic Choice discovers discontinuous constructions and makes them contiguous, it introduces new dictionary terms which did not occur in the text (*wuerde dies geschehen, so* becomes *falls dies wuerde geschehen, so*), it deletes words and non-terminal constituents which can be predicted (like punctuation marks) or which become features of other words (like prefixes, which become a feature of their verb). The power of Syntactic Choice also permits the linguist to make a decision among competing (ambiguous) analyses on the basis of linguistic or pragmatic considerations.

Syntactic Choice uses information contained in the syntactic analysis rules subsequently called Choice Statement, and information which is stated in the separate Syntactic Choice grammar.

Syntactic Choice is performed in three phases. During the first phase all parasitic interpretations, i.e., those interpretations not dominated by S, are deleted. If two or more S readings dominate a text span, the S dominating the longest span is retained.

During the second phase, all Choice Statements contained in the syntactic rule and all choice rules called by the applied syntactic rules, are executed.

During the third phase, Syntactic Choice prints out the set of tentative standard sub-trees in what is called the "parallel work space format". This format permits the standard analysis algorithms and transfer analysis algorithms to analyze a constituent only once, even if it occurs n times in n standard strings, i.e., if the sentence had been n times ambiguous.

As can be seen from the function of the Syntactic Choice component, the Syntactic Choice grammar is easily the most important grammar of every language dealt with in LRS. The combination of choice statements associated with syntactic rules and the choice rules called by them are in effect an algorithm designed by the linguist to compute the underlying structure or structures from which the given surface structure may have been derived. As has been stated above, the computed underlying standard structures are tentative, even more so if the sentence analyzed contains a lexical collocation. It is for this reason that a final check performed by standard analysis grammar is necessary. A detailed description of the decisions performed by the choice instructions and choice rules is given in Appendix II.

The basic deficiency of the Choice Algorithm is that choice rules cannot call other choice rules. Admittedly, to permit unrestricted recursive application of choice rules would probably cause a large number of, if not all, the problems that have been described for the application of reverse transformation rules with transformational grammars. However, the depth of recursion could be restricted to one, while separating, for example, the choice grammar rules into "mother" rules and "daughter" rules. These would be such that the mother rule would call any number of daughter rules at a time but that no daughter rule could call a mother rule or another daughter rule. This will facilitate the construction of the German choice grammar considerably, as well as reduce both the amount of information contained in each rule and its size. To give an example: the choice rules with the names V AC:VSO, V AC:VSOA, V AC:VSO2A, V AC:VSO3A, V AC:VSO4A interpret clause structures consisting of a surface subject, predicate, surface object, with zero through four adverbials, respectively. The letter combination AC: stands for "dealing with a clause structure in which the predicate appears superficially as a verb in the active voice". In each of these rules, a basic type of decision is made again and again. For example:

1. Is the verb indeed in the active voice? This question is decided by testing whether the surface object is a reflexive pronoun and whether the surface subject fulfills the semantic conditions required by the object of the verb. Example: *die Buecher verkaufen sich gut = the books sell well.*

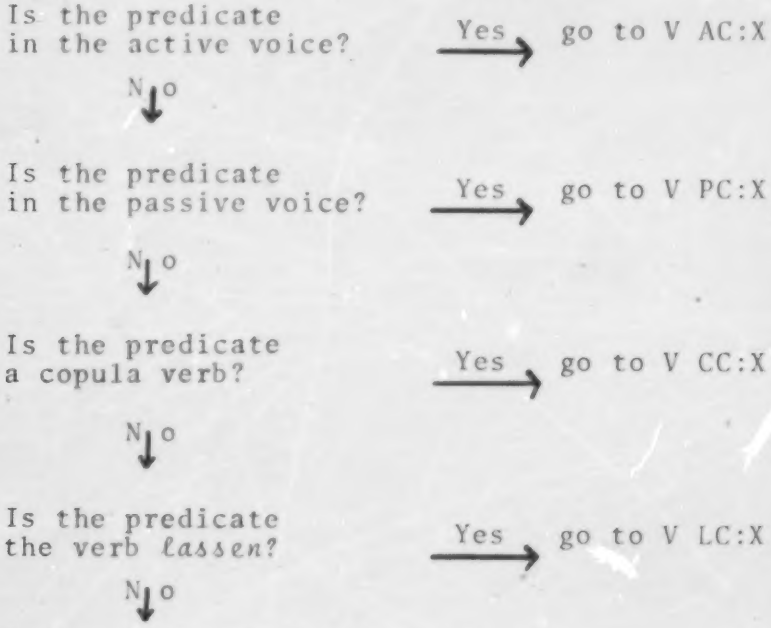
2. Is the verb really an intransitive verb? *Er benimmit sich schlecht = he behaves badly.*

3. Is the verb transitive with one object complement, and which consituent is the object complement— the noun phrase (O) or one of the adverbials (A)? *Er sah Fritz im Garten* vs. *Er befasste sich mit Musik.*

Not only was it necessary to repeat these decisions and numerous others again and again for each choice rule (which interpreted clauses), but the same decisions had to be repeated within the same choice rule but involving different adverbials. For example, the question "are we really dealing with a transitive verb that requires two object complements"? would be asked in V AC:SVO3A for O and the first A, for O and the second A, and for O and the third A. Each test involves the surface predicate, the surface object, and the two arguments being tested as possible object complements.

The amount of information thus accumulated in each choice rule exceeded the amount of allotted core storage. It was therefore necessary to restrict the choice statements in each choice rule to the structures occurring in the 42 sentences analyzed.

Another limitation was that decisions that had been made in earlier choice rules, as for example grammatical relations which had been recognized, were not directly available for later choice rules. Such access to previous results could be simulated by asking in a choice statement of a syntactic rule whether a particular choice statement of a particular choice rule had been successful or not. Depending on the answer, the flow of the choice algorithm could be directed to a choice rule whose name reflected the answer to the question. Thus the selection of the appropriate clause choice rule was determined as represented in the following graph.



AC stands for "active clause", PC, CC, and LC for "passive clause", "copula clause" and "*lassen* clause", respectively. X here represents an acronym consisting of the first letters of the assumed functions of the surface constituents. A surface structure consisting of a noun phrase in the nominative, predicate (represented either as V PRED or V MODAUX... V VERBAL or V PRED... V PRFX), a noun phrase and two adverbials could potentially call the clause rules V AC:SVO2A, V PC:V2O2A (in a passive clause the surface subject is an underlying object), V CC:SO2A (in copula clauses the predicate information was not conveyed to the choice rule), V LC:SVO2A.

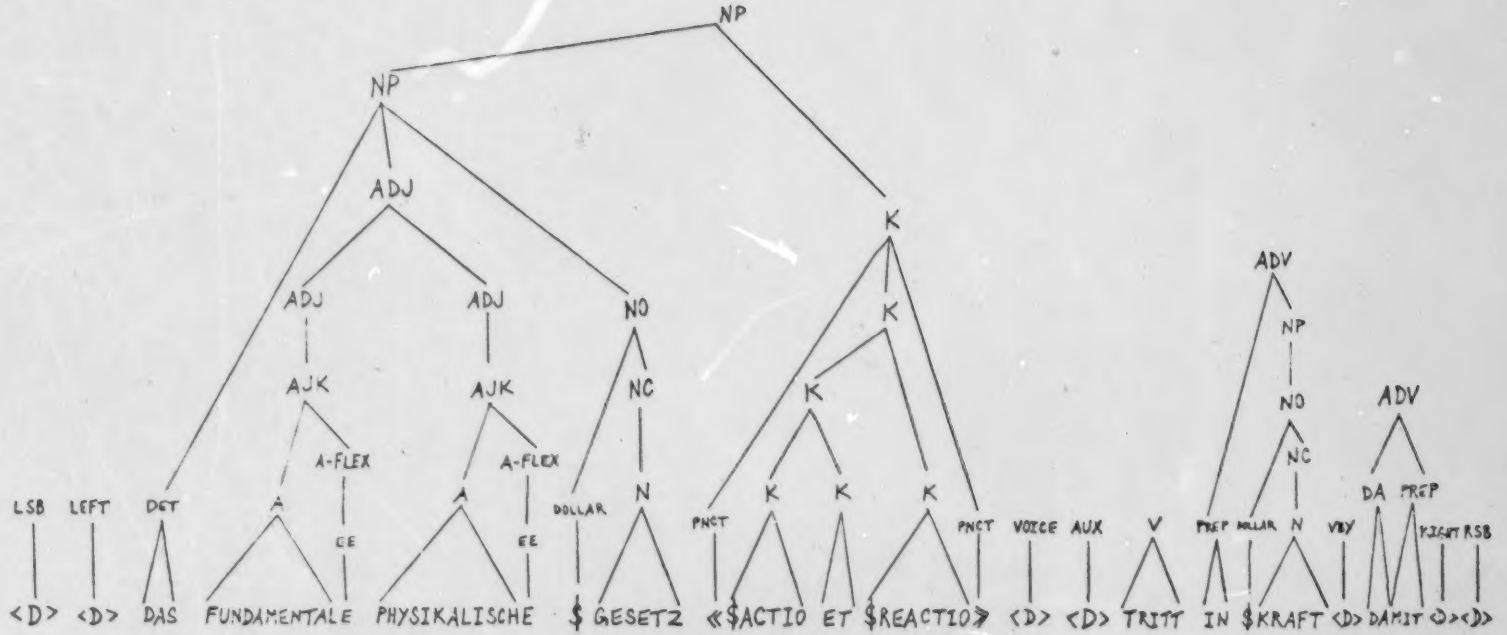
The effect of these restrictions was that the number of choice rules necessary became very large, as did the amount of information to be processed in each rule.

[For Fig. II.8 refer to following page]

Fig. II.8

S 16

SYNTACTIC CHOICE



2.4 The Standard Components

It is the function of the Standard components to derive the standard reading of a sentence (or readings, in cases of ambiguity) which can subsequently be interpreted by the meaning assignment grammar or normal form grammar. The input for the Standard components are the set of tentative standard trees provided by the syntactic choice component. Its output are a set of well-formed standard trees, that is, trees which are dominated by the symbol S. If a sentence was structurally ambiguous and no selection was made during Syntactic Choice more than one standard reading will be generated. (Lexical ambiguity does not result in multiple standard analyses.) Standard Analysis is performed by three components, the lexical collocation component, the standard syntactic component, and the standard choice component.

[For Fig. II.9, refer to following page]

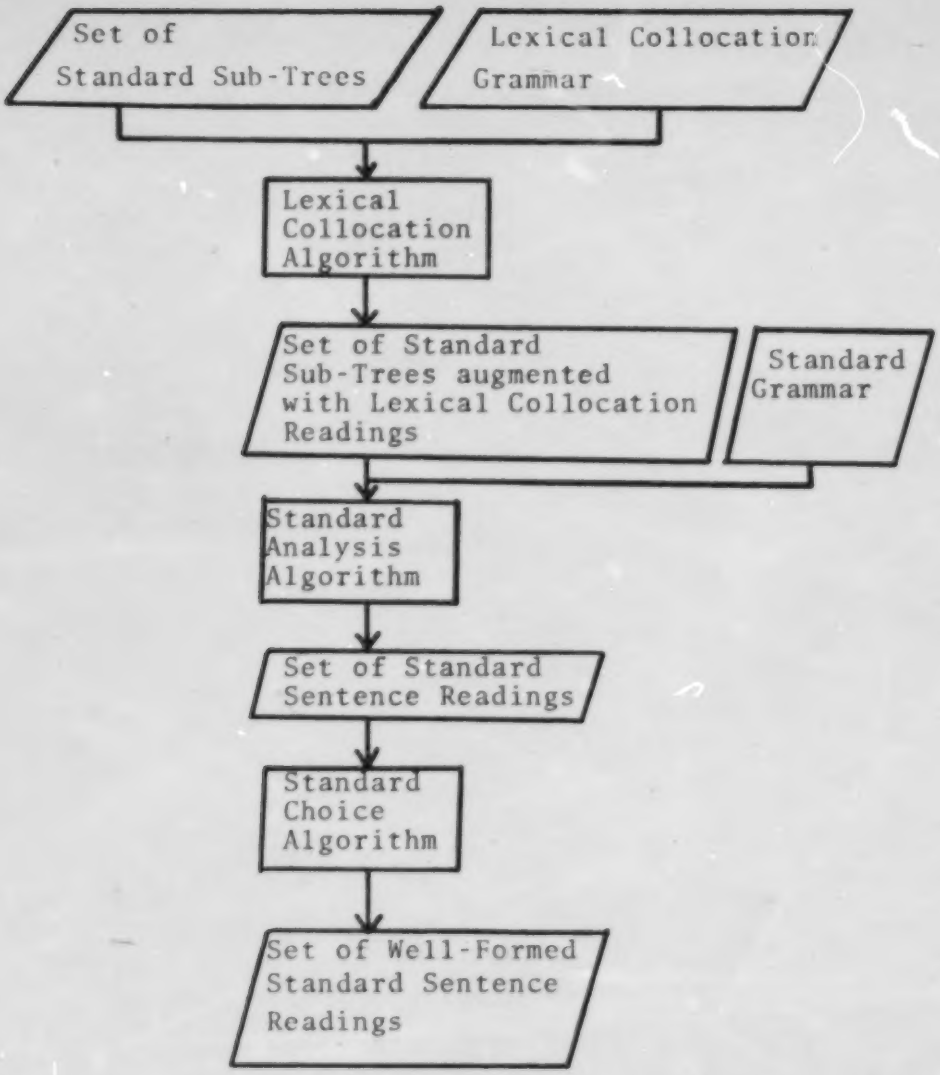

Set of Standard Sub-Trees
Lexical Collocation Grammar
Lexical Collocation Algorithm
Set of Standard Sub-Trees augmented with Lexical Collocation Readings
Standard Grammar
Standard Analysis Algorithm
Set of Standard Sentence Readings
Standard Choice Algorithm
Set of Well-Formed Standard Sentence Readings

Figure II.9

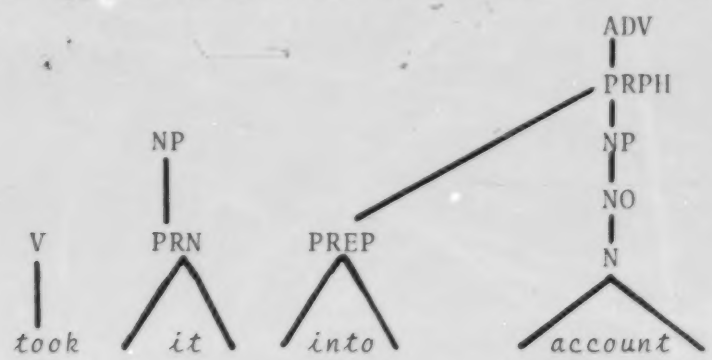
2.4.1 The Lexical Collocation Component

It is the purpose of Lexical Collocation analysis to find all sequences of terminal symbols which form what we call a lexical collocation, without affecting the structural interpretations of the individual components since these might be needed to determine possible transformations.

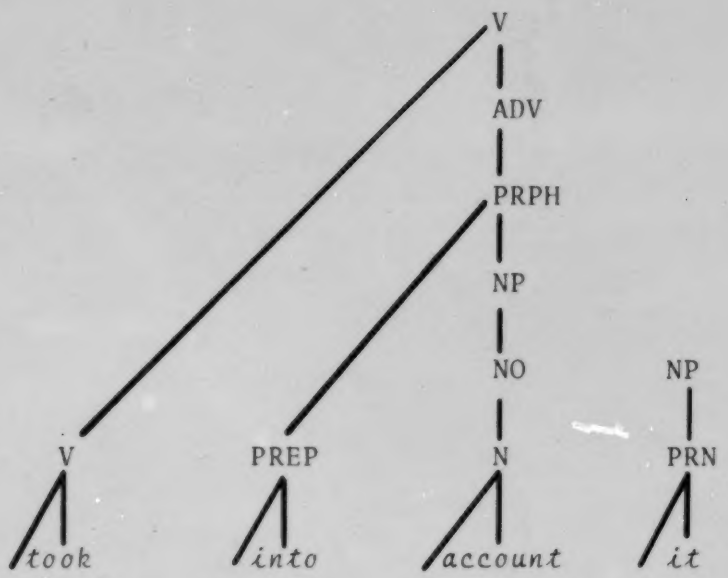
In a surface text, elements of a lexical collocation can occur in any order and at any distance, that is, they may occur discontinously. After syntactic choice, however, we can determine precisely, if a lexical collocation occurs, which order the individual elements will occur in, their syntactic superstructure, and which constituents, if any, occur between them, obligatorily and optionally. Consider the rule

```
V V      C 145    V NP    V ADV    C 100.8    C 23000
+                         O 1                 V ADV
+                                             D 5,6
.
.
.
R 2,6
S 2-6-3-4
```

which interprets *take into account*. This rule is to be read as: "The sequence of terminals C 100.8 (*into*) and C 23000 (*account*), which is dominated by the constituent ADV covering the text span from the 5th to the 6th rule term, may form a lexical collocation with the dictionary entry C 145 (take, took, taken). Between these terminals, adverbials may optionally occur (signified by the operator O below V ADV). The maximum number of adverbials which may occur in this position is 1 (O 1). A noun phrase must occur between the terminal *take* and the sequence *into account*. The rule antecedent V V (verb) rewrites only the second and sixth rule term; the rule terms in the rule consequent are to be arranged in the order 2, 6, 3, 4 (the rule antecedent is always counted as rule term 1). The application of the rule above to the structure



would thus result in the structure



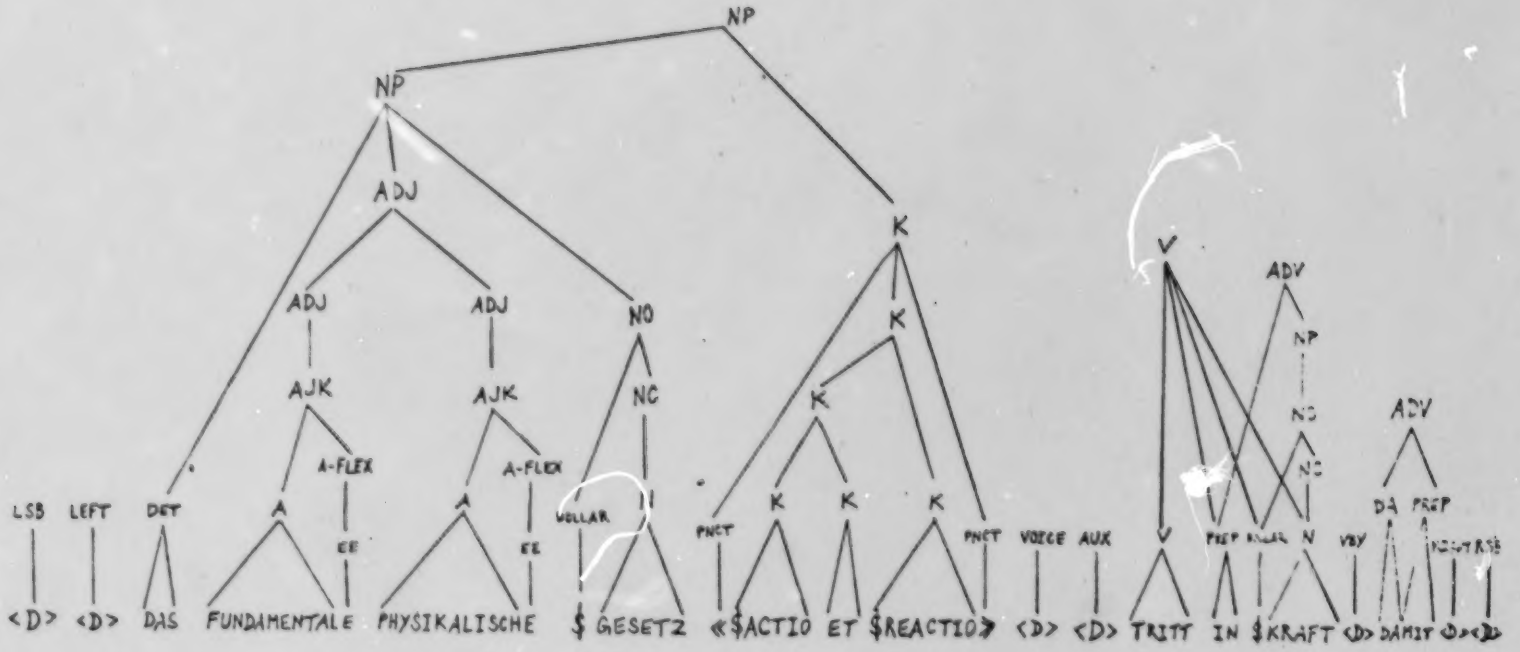
As described in Section I above, time and cut in funding with corresponding reduction of programming personnel did not permit the construction of a lexical collocation component as envisioned. The current lexical collocation component is basically identical to the word analysis component, with the exception that it can process input in the "parallel work space" format, the output of syntactic choice. The lexical analysis component is not sensitive to flags which have been set by syntactic choice. It investigates for every terminal whether it can form a lexical collocation with the terms occurring to its left. Since the word analysis algorithm cannot process optional symbols, a number of lexical collocations had to be coded several times to account for variations like singular or plural *he made up his mind, they made up their minds*, and for each optionally occurring internal variable slot. In view of this undesirable multiplication of entries for lexical collocations, we endeavored to avoid additional proliferation of such entries resulting from stating 0,1,2,3 etc. optional symbols that might occur between the significant terminal elements of a lexical collocation. It was thus necessary to add tests to each clausal choice rule which tested for each indirect object and all adverbials whether it could possibly form a lexical collocation with the main verb. If so, the corresponding element was moved directly to the right of the verb in the resulting standard string. Another difference between the envisioned lexical collocation component and the current one had no effect on grammar design; however, it had an effect on execution time. Originally, lexical collocation analysis was only to be called if elements which could possibly be members of a lexical collocation occurred

in a text. The current version applies lexical collocation analysis obligatorily for all the sentences.

[For Fig. II.10 refer to following page]

Fig. II.10

S 16
LEXICAL COLLOCATION ANALYSIS



2.4.2 The Standard Analysis Component

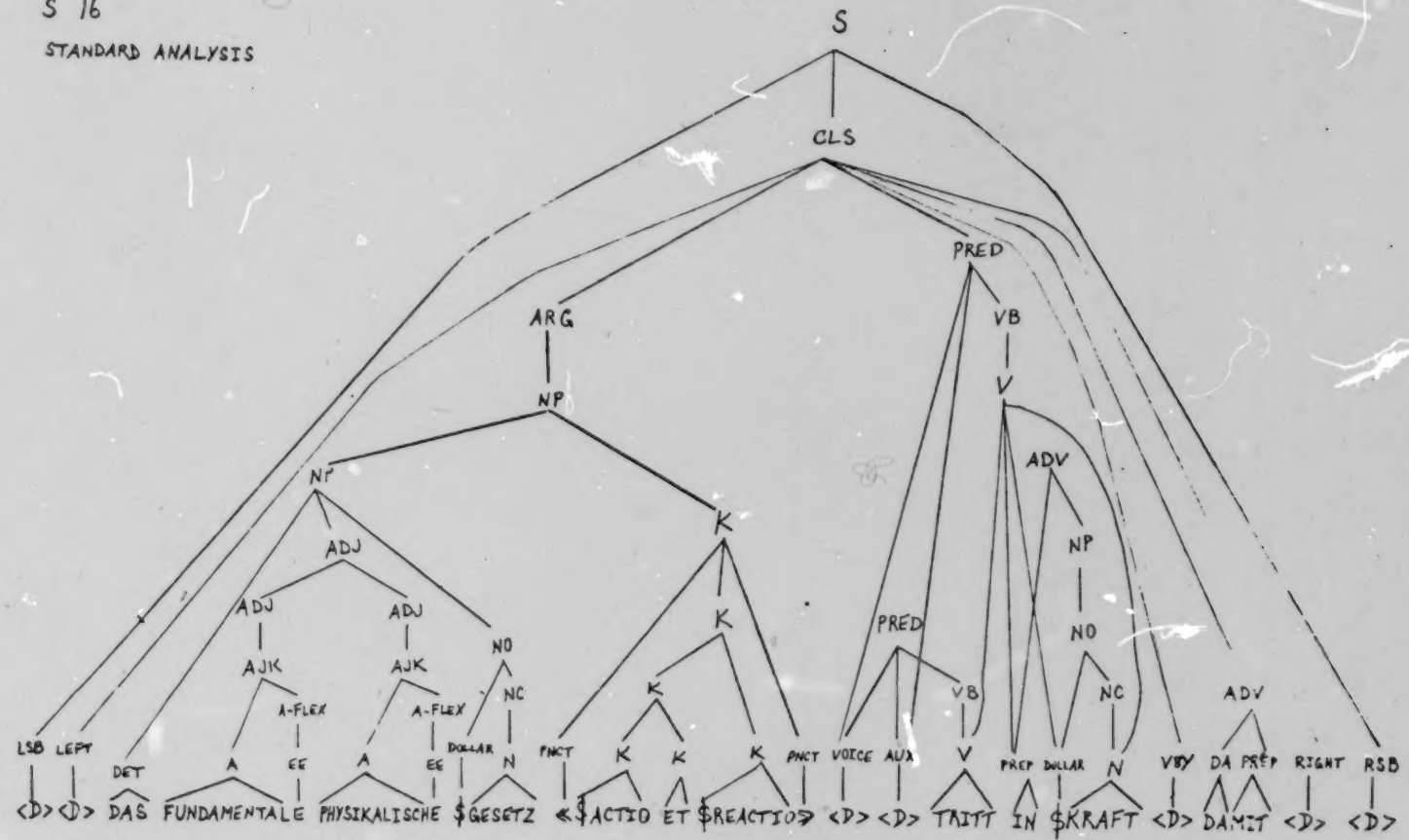
The purpose of the Standard Analysis is to construct a complete syntactic tree dominated by S to provide a complete syntactic description for the sequences of structures introduced and retained by syntactic choice. The Standard Analysis algorithm is identical to the lexical collocation algorithm, but it only operates on constituents which were unflagged by syntactic choice. Lexical collocation is similarly insensitive to flags. Standard Analysis, which operates with the Standard Analysis grammar of a language, is very small in size. The number of clause rules, for example, is 15. It is because of the predefined order in which the constituents occur, that the various syntactic and/or semantic relations that must hold between constituents can be expressed economically by means of a small number of grammar rules. The purpose of standard grammar is basically twofold:

a. it assigns the correct structure to sentences which contain lexical collocations, and

b. it regenerates the syntactic superstructure that was destroyed during syntactic choice.

[For Fig. II.11 refer to following page]

Fig. II.11

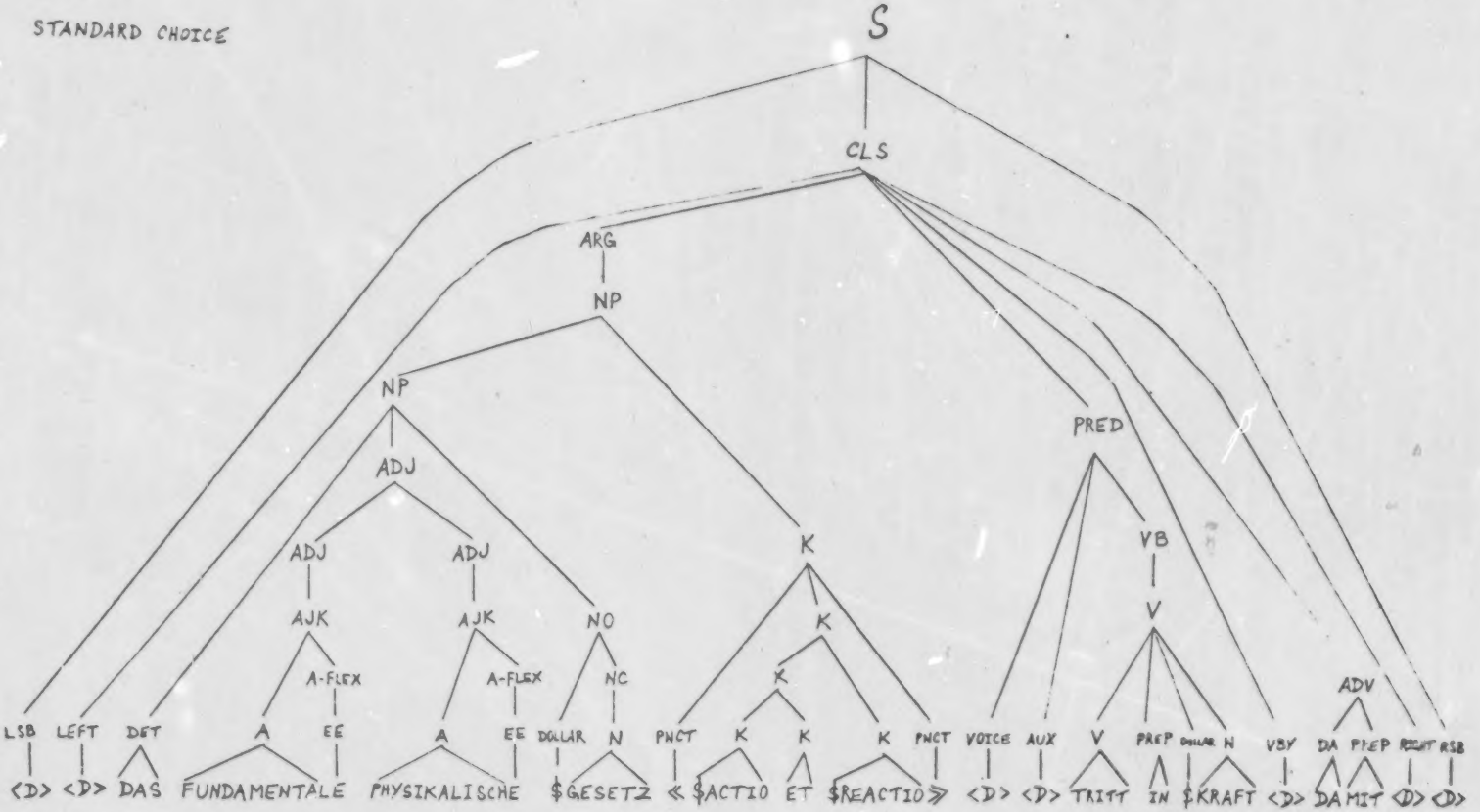

S 16
STANDARD ANALYSIS
S
CLS
ARG
NP
NP
ADJ
ADJ
ADJ
AJK
AJK
A-FLEX
A-FLEX
A
EE
A
EE
NO
NC
DOLLAR
N
K
K
K
K
K
PRED
VB
V
ADV
NP
NO
NC
PNCT
K
K
K
PNCT
VOICE
AUX
V
PREP
DOLLAR
N
VBY
ADV
DA
PREP
RIGHT
RSB
LSB
LEFT
DET
<D>
<D>
DAS
FUNDAMENTALE
PHYSIKALISCHE
$GESETZ
«$ACTIO
ET
$REACTIO»
<D>
<D>
TRITT
IN
$KRAFT
<D>
DA.MIT
<D>
<D>

2.4.3 The Standard Choice Component

The purpose of the Standard Choice component is to delete all readings which are not well-formed according to the standard grammar, i.e., not dominated by the symbol S. It is similar in operation to the syntactic choice component, with the difference that it accepts a work space in parallel format. Standard Choice does not introduce additional dummies nor does it delete rule terms. It does perform additional disambiguation and permits the rejection of whole trees if they compete with trees which dominate a lexical collocation with the feature LX(P) for "lexical collocation with preference". The Standard Choice component does not operate with a separate standard choice grammar. All instructions executed by Standard Choice are part of the standard rules. Choice statements which are attached to structures which were already constructed during syntactic analysis and consequently executed by syntactic choice are not repeated by Standard Choice.

[For Fig. II.12 refer to following page]

Fig. II.12


S 16
STANDARD CHOICE
S
CLS
ARG
NP
NP
ADJ
ADJ
ADJ
AJK
AJK
A-FLEX
A-FLEX
A
EE
A
EE
NO
NC
DOLLAR
N
K
K
K
K
K
PNCT
PNCT
PRED
VB
V
ADV
VOICE
AUX
V
PREP
DOLLAR
N
VBY
DA
PREP
RIGHT
RSB
LSB
LEFT
DET
<D>
<D>
DAS
FUNDAMENTALE
PHYSIKALISCHE
$GESETZ
≪
$ACTIO
ET
$REACTIO
≫
<D>
<D>
TRITT
IN
$KRAFT
<D>
DA
MIT
<D>
<D>

2.4.4 The Transfer Components

The purpose of the Transfer Analysis is to associate with each word in a standard reading its word sense— if the word was disambiguated— or its word senses if the word is still ambiguous. Similarly, it associates with each sub-tree or with connected sub-trees (mother, daughters, and possibly daughters and descendents of daughters) a functional interpretation. These interpretations are referred to as normal form expressions. The Transfer Analysis component operates with the transfer grammar of a language and determines for each standard text the normal form readings applicable to that text. Transfer choice eliminates all normal form readings which do not take part in a contiguous coverage of the standard tree structure. Again, a special P-operator attached to a transfer rule permits transfer choice to select such readings in preference to other transfer readings covering the same structure.

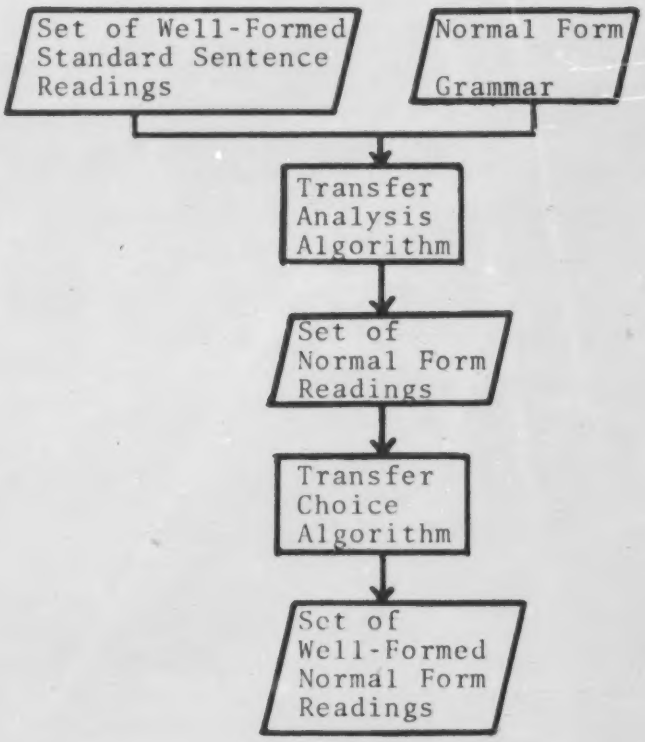

Set of Well-Formed Standard Sentence Readings
Normal Form Grammar
Transfer Analysis Algorithm
Set of Normal Form Readings
Transfer Choice Algorithm
Set of Well-Formed Normal Form Readings

Fig. II.13

The actual design of the Transfer Analysis and choice algorithm generally corresponds to the design described in the previous reports. Their description can be found in Appendix III. However, there was not time to add one essential feature. The current Transfer Analysis algorithm does not handle K-spans which have been recognized during dictionary and word analysis (cf. Section 2.3.1 above). For this reason all such spans like *Newton* and *actio, et, reactio* had to be entered into the transfer dictionary to permit their translation.

[For Figs. II.14 and II.15 refer to following page]

Fig. II.14

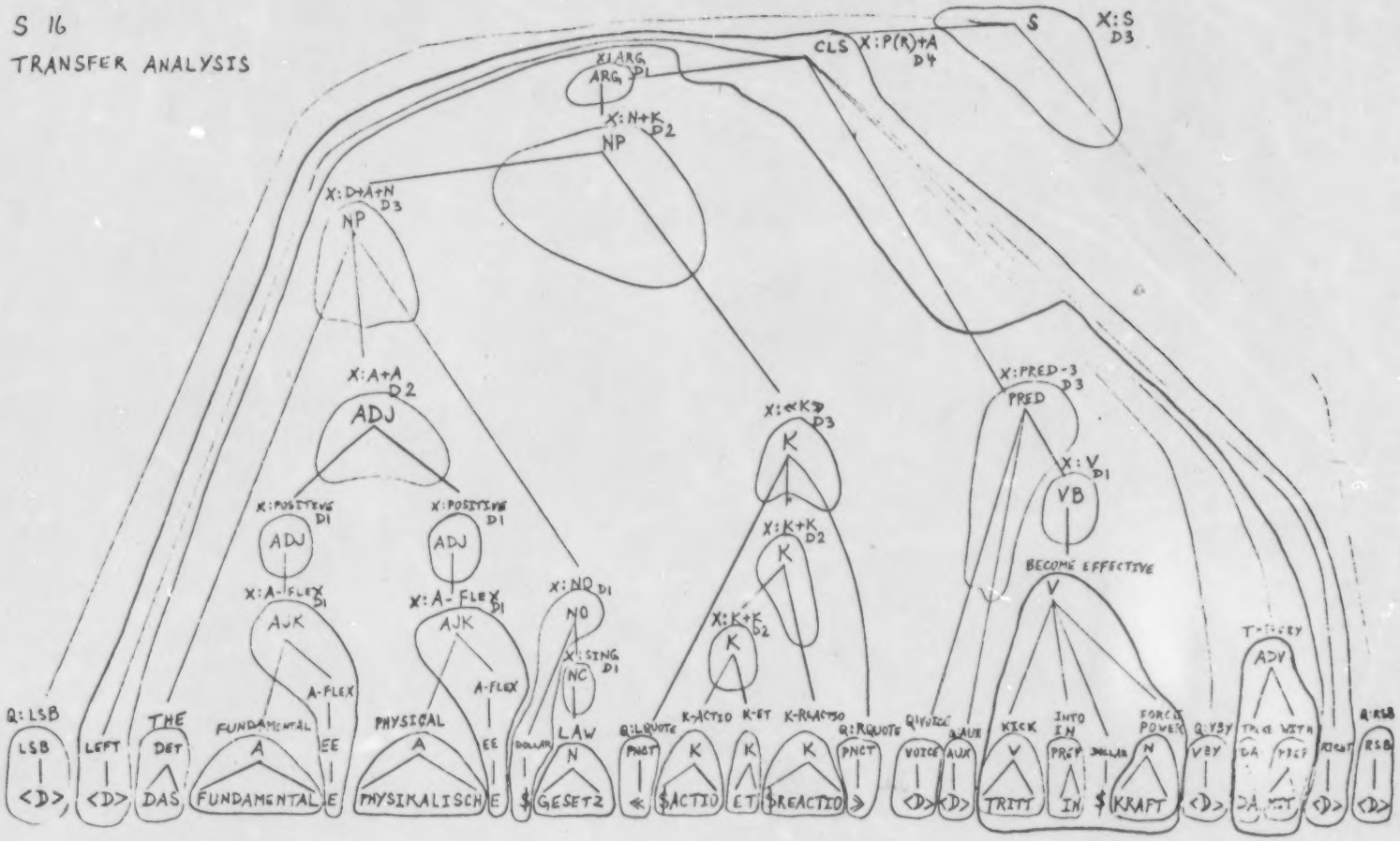
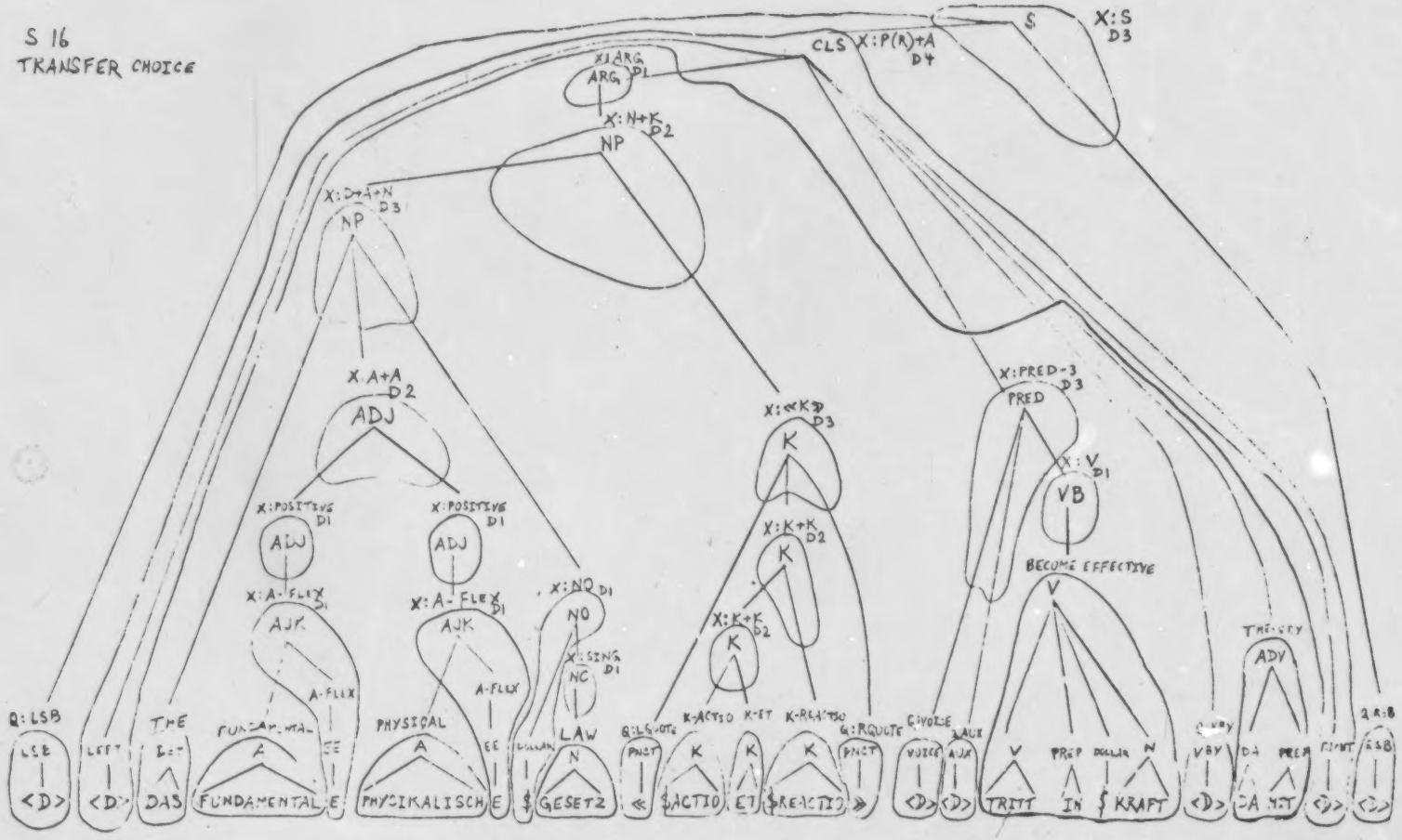

S 16
TRANSFER ANALYSIS
S
X:S D3
CLS X:P(R)+A D4
X1ARG ARG D1
X:N+K D2
NP
X:D+A+N D3
NP
X:A+A D2
ADJ
X:POSITIVE D1
ADJ
X:A-FLEX D1
AJK
A-FLEX
EE
FUNDAMENTAL
A
DAS
THE
DET
LEFT
LSB
Q:LSB
<D>
X:POSITIVE D1
ADJ
X:A-FLEX D1
AJK
A-FLEX
EE
PHYSICAL
A
X:NO D1
NO
X:SING D1
NC
DOLLAR
LAW
N
$
GESETZ
X:K+K D3
K
X:K+K D2
K
X:K+K D2
K
Q:LQUOTE
PNCT
«
K-ACTIO
K
$ACTIO
K-ET
K
ET
K-REACTIO
K
$REACTIO
Q:RQUOTE
PNCT
»
X:PRED-3 D3
PRED
X:V D1
VB
BECOME EFFECTIVE
V
Q:VOICE
VOICE
<D>
Q:AUX
AUX
<D>
KICK
V
TRITT
INTO
IN
PREP
IN
DOLLAR
$
FORCE POWER
N
KRAFT
Q:VBY
VBY
<D>
ADV
DA
PREP
Q:RSB
RSB
<D>

Fig. II.15


S 16
TRANSFER CHOICE
S
X:S
D3
CLS X:P(R)+A
D4
X:ARG
ARG
D1
X:N+K
D2
NP
X:D+A+N
D3
NP
X:A+A
D2
ADJ
X:POSITIVE
D1
ADJ
X:A-FLEX
D1
AJK
A-FLEX
X:POSITIVE
D1
ADJ
X:A-FLEX
D1
AJK
A-FLEX
X:NQ
D1
NQ
X:SING
D1
NC
X:«K»
D3
K
X:K+K
D2
K
X:K+K
D2
K
X:PRED-3
D3
PRED
X:V
D1
VB
BECOME EFFECTIVE
V
Q:LSB
LSB
<D>
LEFT
<D>
THE
DAS
FUNDAMENTA-
A
FUNDAMENTAL
E
PHYSICAL
A
PHYCIKALISCH
E
$
LAW
N
GESETZ
Q:LGUOTE
PNCT
«
K-ACTIO
K
$ACTIO
ET
K-ET
K
K-REACTIO
$REACTIO
Q:RQUOTE
PNCT
»
Q:VOICE
VOICE
<D>
AUX
<D>
V
TRITT
PREP
IN
DOLLAR
$
N
KRAFT
Q:VBY
VBY
<D>
THEORY
ADV
D4
PREP
-A
NT
PUNCT
<D>
2 R:B
RSB
<D>

2.5 The Standard Production Components

The purpose of Standard Production is to associate with the normal form reading of a sentence the standard trees of all English output sentences having the same normal form reading. In other words, we try to find all English sentences which have the same meaning as the German sentence. As a second function, Standard Production tries to arrange the constituents of such sentences in an order which closely reflects an acceptable English surface order. The Standard Production component consists of two sub-components:

a. file entry construction

and

b. output standard choice.

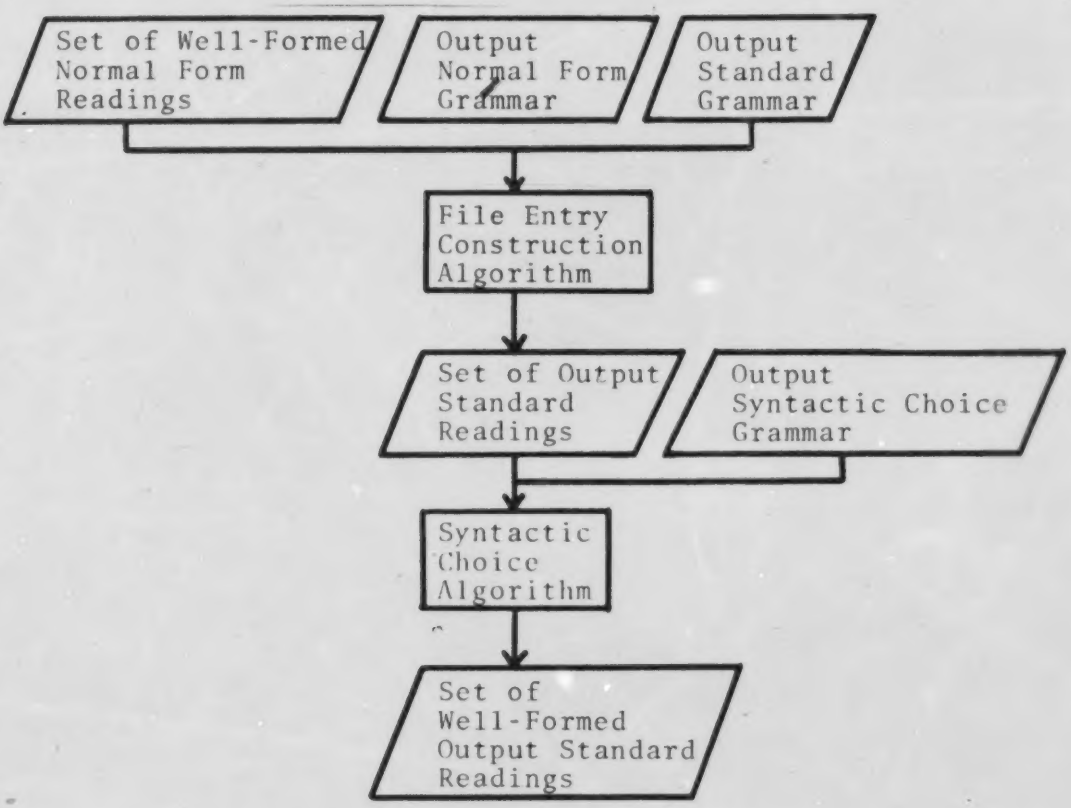

Set of Well-Formed Normal Form Readings
Output Normal Form Grammar
Output Standard Grammar
File Entry Construction Algorithm
Set of Output Standard Readings
Output Syntactic Choice Grammar
Syntactic Choice Algorithm
Set of Well-Formed Output Standard Readings

Fig. II.16

2.5.1 The File Entry Construction Component

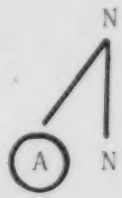
The purpose of File Entry Construction is to associate with the normal form readings provided by the analysis of a German input sentence all well-formed English standard trees (dominated by the symbol S) with the same meaning as the input sentence. At the same time, File Entry Construction tries to produce translation sentences whose structural description is similar to the description of the German sentence, in order to reduce the number of translations provided for each input sentence.

File Entry Construction operates with the English transfer grammar and the English standard grammar. It is executed in two phases, Interlingual Mapping, and Synthesis.

2.5.1.1 Interlingual Mapping

The purpose of Interlingual Mapping (the name does not quite correctly reflect its function) is to associate with the German normal form expression all English sub-trees interpreted by the corresponding English normal form expressions. At the same time, it checks whether the retrieved standard sub-trees can be connected. Two standard sub-trees can be connected if the label of the root of a sub-tree A is identical to a label of a node of sub-tree B into which A is to be linked. Thus the sub-trees

and



can be connected in

. The sub-trees

and

cannot be connected.

File Entry Construction constructs the superficial English standard trees by working from left-to-right and bottom-to-top. It performs, essentially, the following five operations.

1. Retrieval of the English transfer rules with the same normal form expression;

2. performance of subscript/value check (a German normal form expression is interpreted as a rule condition which must be satisfied by the English normal form expression). The English normal form expressions which do not match are deleted;

3. retrieval of all English standard sub-trees interpreted by the matching English normal form expressions;

4. conflation of terminal symbols with identical category symbols;

5. connection of the retrieved standard sub-trees with the sub-trees obtained so far (recall that the algorithm works from bottom-to-top) and rejection of those sub-trees which do not connect.

2.5.1.2 Synthesis

The function of the Synthesis portion of File Entry Construction is to check that the superficial standard trees are indeed well-formed according to the standard grammar of the language. Those sub-trees which do not satisfy this condition are deleted. At the same time, Synthesis checks that nodes of the remaining trees have the subscripts and values which were stated in the English transfer rules.

Synthesis operates with the English standard rules that have been provided by interlingual mapping. For each rule it basically performs four operations:

a. subscript/value check for each term of the rule;

b. operations;

c. left-side construction, and

d. left-side check.

During left-side check, the constructed left side is compared against the left side stipulated by the English transfer rule. Only those columns of the constructed left side which contain values which match the values provided by transfer are retained. For example, the English transfer rule V SING (singular) refers to a sub-tree which builds on a noun and noun endings, and specifies that the result should be the number singular. The result of left-side construction, however, is the subscript NU with the values singular and plural. Left-side check eliminates the column or columns containing the value plural.

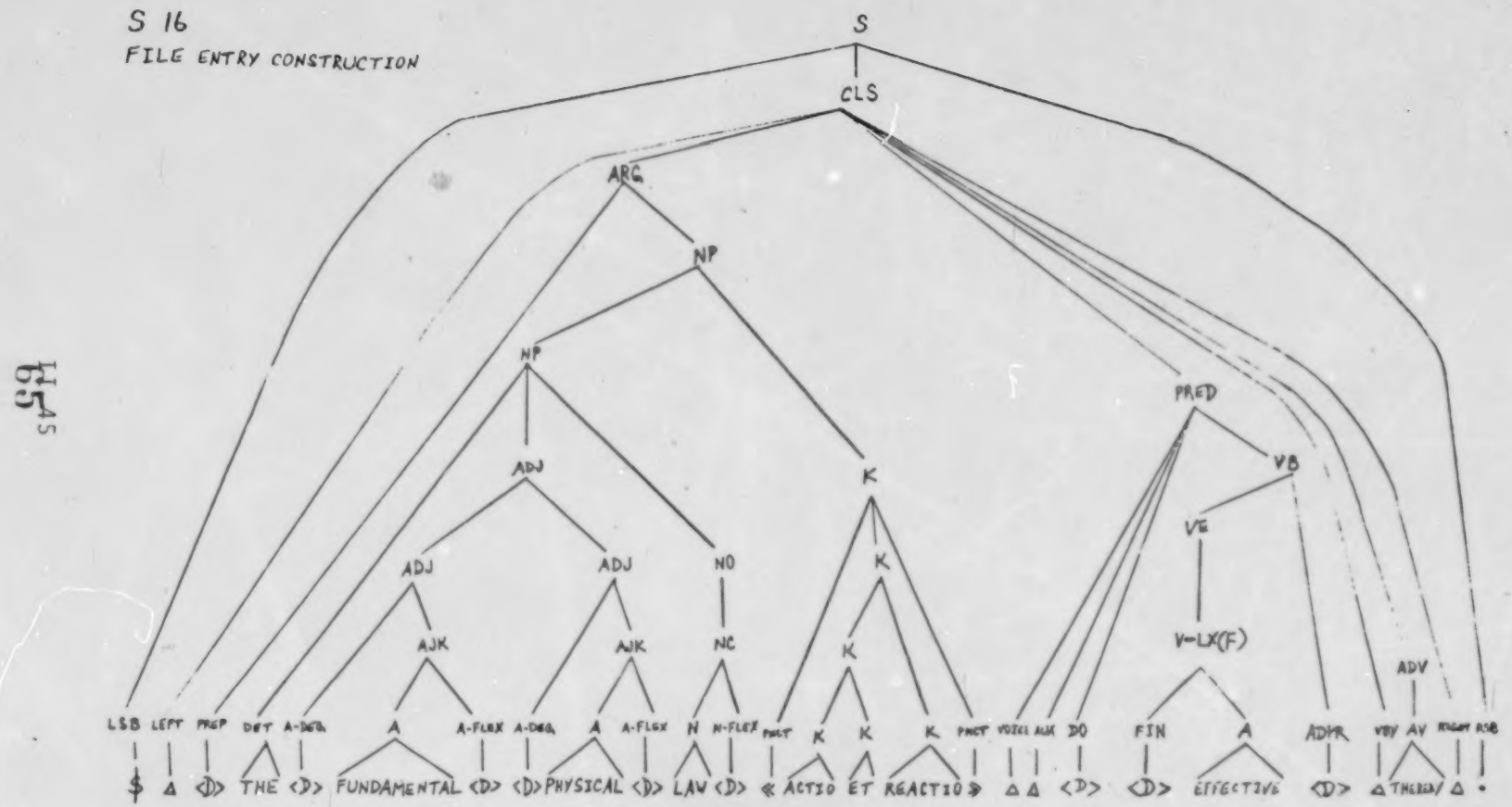
The output of synthesis is a well-formed English standard tree (or set of trees in cases of structural ambiguity of the input text).

2.5.1.3 Standard Output Text

We mentioned above that sometime during the work on the experiment we felt that the printout of the standard string, as we had originally envisioned it (with canonized endings like CITY-S, or preceding endings like -S WATCH), might not be convincing examples of the system's capability of performing quality translation of individual sentences— at least, not convincing to the non-specialist in the field. We therefore decided to modify the choice component and lexical spellout component of the rearrangement phase and to operate on the output of File Entry Construction. The aim was to produce English output which looks like normal English text, the only exception being that the inflectional endings of verbs in lexical collocations appear in front of the lexical collocation. At the same time we decided to provide a printout format in which different translations of words would be printed below one another. The problem which confronted us here was to guarantee that co-occurrence relations between words were not disturbed by this arrangement. Example: we had to make sure that *more* occurred with *extensive*, but *er* with *great*, that *consider* occurred with zero preposition, but *think* with *of*; that nouns and verbs preceded their proper endings: *men* with zero ending but *boy* with the ending *s*; *stop* with the past ending *ped* vs. *terminate* with *ed*; that *writer* had to be preceeded by *a* and *author* by *an*. Since such co-occurrence relations also hold between non-contiguous words, and in order to facilitate the construction of the lexical spellout component, we decided to associate with a constituent, if necessary, a subscript whose value referred to the location of the terminals which had to co-occur. This fact, and the fact that File Entry Construction did not perform conflation of non-terminal nodes, had a decisive effect on the English standard grammar description. The English adjective, verb, and noun morphology were most affected; we had to introduce explicit zero morphemes (which were represented by blanks). These were not to be printed by lexical spellout. Zero morphemes did not only occur under endings, but also under prepositions (consider the zero preposition in *he considers John*, as opposed to *he thinks of John*), auxiliaries and modals, and others (for details, cf. Standard Grammar in Appendix II).

[For Fig. II.17 refer to following page]

Fig. II.17


S 16
FILE ENTRY CONSTRUCTION
S
CLS
ARG
NP
NP
NP
ADJ
ADJ
ADJ
AJK
AJK
NO
NC
K
K
K
K
PRED
VB
VE
V-LX(F)
ADV
LSB LEFT PREP DET A-DEG A A-FLEX A-DEG A A-FLEX N N-FLEX PNCT K K K PNCT VOICE AUX DO FIN A ADVR VBV AV RIGHT RSB
$ Δ ◁D▷ THE ◁D▷ FUNDAMENTAL ◁D▷ ◁D▷ PHYSICAL ◁D▷ LAW ◁D▷ « ACTIO ET REACTIO » Δ Δ ◁D▷ ◁D▷ EFFECTIVE ◁D▷ Δ THERE/ Δ .

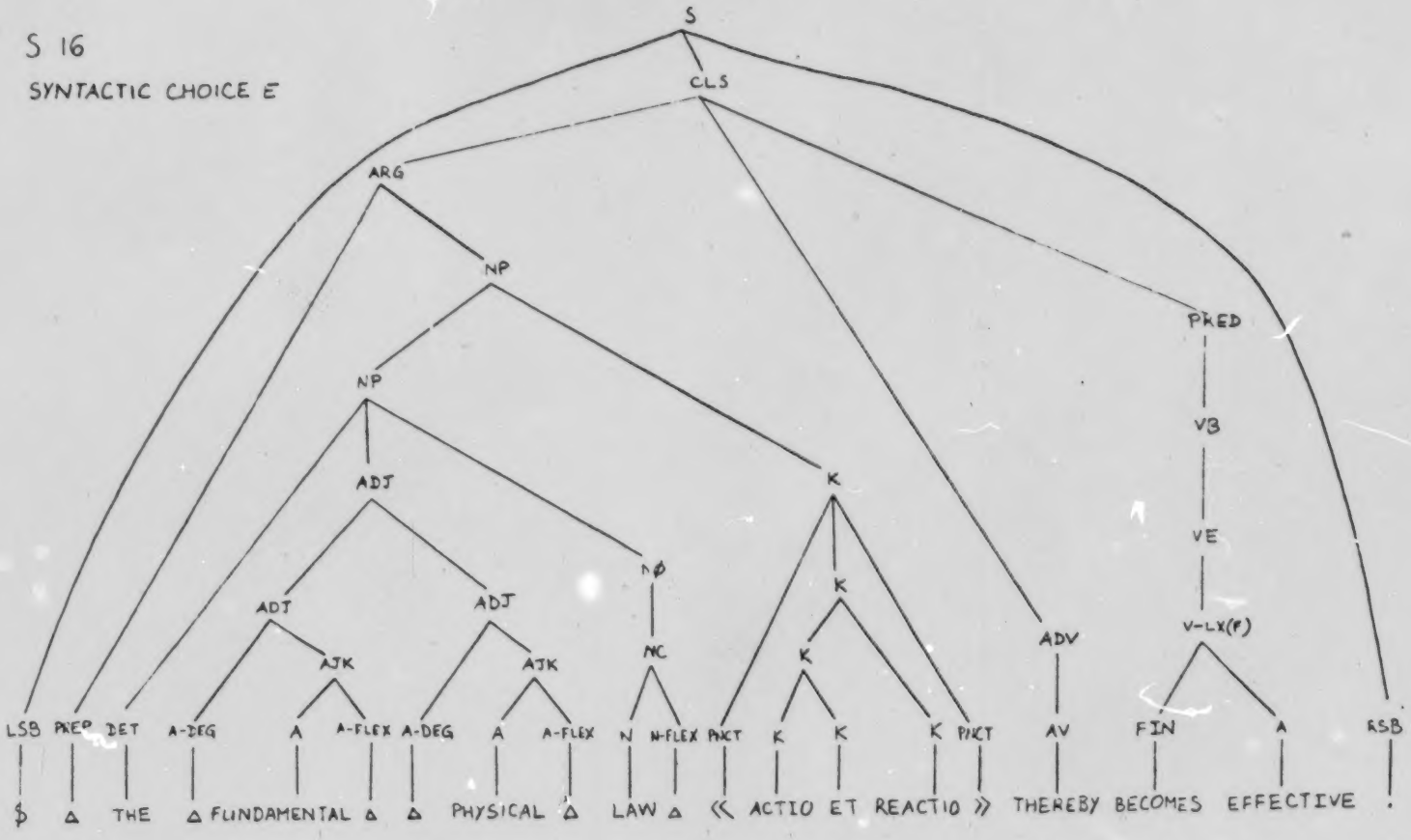
2.5.2 Output Syntactic Choice

The purpose of Output Choice is basically twofold. It selects the values of lexical dummies that need to be printed by lexical spellout, and it determines the order in which the constituents are to be printed out. Output Syntactic Choice uses the English standard choice grammar as well as the choice instructions contained in the English standard rules.

The output standard choice algorithm is nearly identical to the first two subcomponents of the Syntactic Choice algorithm, with the following differences. It can accept a work space in parallel format; it does not destroy any nodes. (English standard dummy terms, all of which are rewritten as [DUMMY], are not printed by the lexical spellout algorithm.) It can thus perform all operations of Syntactic Choice with the difference that permutations are only indicated rather than executed.

[For Fig. II.18 refer to following page]

Fig. II.18


S 16
SYNTACTIC CHOICE E
S
CLS
ARG
NP
NP
NP
ADJ
ADJ
ADJ
NØ
NC
K
K
K
K
K
PRED
VB
VE
V-LX(F)
ADV
LSB
PREP
DET
A-DEG
A
A-FLEX
A-DEG
A
A-FLEX
N
N-FLEX
PNCT
K
K
K
PNCT
AV
FIN
A
RSB
$
Δ
THE
Δ FUNDAMENTAL Δ
Δ
PHYSICAL Δ
LAW Δ
« ACTIO ET REACTIO »
THEREBY BECOMES
EFFECTIVE
.

2.6 The Lexical Spellout Component

The Lexical Spellout component accepts the output of syntactic choice and produces from it English written text (cf. Fig. II.19). If the text contains multiple translations for lexical pieces, they are printed out in a vertical format. Lexical elements among which a co-occurrence relation holds receive the same numerical subscript. Lexical Spellout does not print DUMMY terminals nor zero morphemes. If an ending is to be printed which has the operator L (for "letter") attached, the last letter of the preceding terminal is repeated. Thus, forms like *bigger*, *stopped* and others are generated.

It may be necessary to point out that the multiple translations provided are a sub-set selected by LRS of the translations listed for a term in the dictionary. The multiple translations are all equally good from the LRS point of view. The system in effect says "I am providing you with multiple translations because either the dictionary does not have a finer classification for me to make a decision, or there is not enough disambiguating information contained in this sentence to permit me to make a finer selection". Obviously, the output format could be changed to show multiple translations in the margin or as footnotes. We feel, however, that we may not supress such translations because they might help the user in attaining a correct interpretation of the meaning of the sentence.

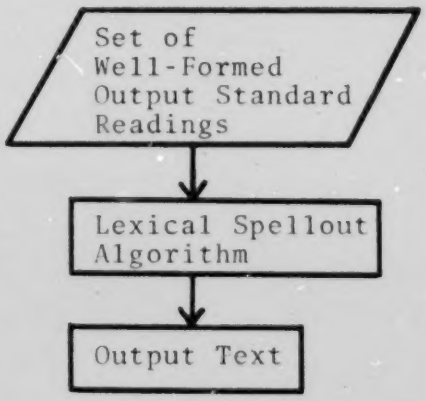


Fig. II.19

[For Fig. II.20 refer to following page]

Fig. II.20

S 16

LEXICAL SPELLOUT

$ THE FUNDAMENTAL PHYSICAL LAW «ACTIO ET REACTIO» THEREBY BECOMES EFFECTIVE.

SECTION III

3. Lexicographic Work Within the Experiment

3.0 Introductory Remarks

The purpose of the lexicographic work done in the experiment described above was to test the Center's established classification system of lexical features in the analysis and translation into English of a sizeable technical text in German. The feature system comprises morphological, syntactic, and semo-syntactic features and selection restrictions.

This section describes the methods employed in the lexicographic work, the features assigned to the lexical entries, the problems we encountered, and the solutions we found.

3.1 Methodology

The lexicographic work was done in two stages: first, the provision of correct English translation equivalents for each word occurring in the German text, and second, the coding of the necessary dictionary rules which enabled the system to analyze the German input text and to select the particular translation automatically.

3.1.1 Provision of Translation Equivalents

To facilitate the provision of English translation equivalents, the German input text was displayed in a concordance format which aligns the concorded text word in the center of the display page and, on the left half of the page, prints the entire sentence in which each word occurs. The right half is left blank. Figure III.1 shows a page of this concordance. Linguists fluent in both German and English examined each concorded word (i.e., each occurrence of each word in the text) separately within its sentence environment and decided on the most appropriate translation for the word in that particular environment. This translation was entered on the blank right side of the concordance page, along with any remarks which seemed relevant to the translator.

[For Fig. III.1 refer to following page]

Fig. III.1

	RWERKERKUNST KANN DIESES ALS DAS WICHTIGSTE SPRAEPARAT WERDEN . +	ANGESEHEN		~~consider~~ ✓	
JSR04314	+ $DIE $BREN NKAMMER BESTAND EBENFALLS AUS EINEM $STAHLZYLINDER MIT $AUSSTROEMDUESE AUS $KUPFER . +	ANGESETZTER	attach	✓	
JSR01508	+ $DIESE $LANZEN KOENNEN ALS GEFIEDERTE $PFEILE WERDEN , WELCHE MIT EINEM $PULVERTREIBSATZ AUSGERUESTET SIND . + ***	ANGESPROCHEN	~~can~~ regarded as...	~~consider~~ ✓	
JSR08707	+ $FUER DAS $WASSERSTOFF -$SAUERSTOFF-$GEMISCH , WELCHES AUCH HEUTE NOCH PRAKTISCH ALS EINES DER ENERGIEREICHSTEN $TREIBSTOFFGEMISCHE	ANGESPROCHEN		regarded as...	
JSR04008	WERDEN MUSS , LAUTET DIE $VERBRENNUNGSGLEICHUNG : + + $MEINE LETZTEN $POSTRAKETEN WURDEN IM $JAHRE 1935 GESTARTET UND WURDEN MIT FLUESSIGEM $TREIBSTOFF . +	ANGETRIEBEN	(1) propel	✓	
JSR02203	+ $EIN $GEBLAESE $G LIEFERT UEBER DEN $VERGASER $C $BRENNSTOFF-$LUFTGEMISCH IN EINE ZUGLEICH ALS $VERTEILER AUSGEBILDETE , EBENFALLS FREMD	ANGETRIEBENE	(1)		
JSR11303	$TURBINE $VSG . + + $NACH LANGEN , MUEHSELIGEN $LABORATORIUMSARBEITEN UEBERRASCHTEN DIESE DREI $PIONIERE AM 11 . $APRIL 1928 PLOETZLICH DIE $WELT MIT EINER GANZ NEUEN $MASCHINE , DEN VON $PULVERRAKETEN $BRENNWAGEN . + ***	ANGETRIEBENEN	(1)		
JSR06905	$WIE ERWAEHNT , BESTEHT DIE IN DER $PULVERHERSTELLUNG $KOHLE NICHT AUS REINEM $KOHLENSTOFF , SONDERN SIE WEIST NOCH GEWISSE IN DER VERKOHLTEN $HOLZ ENTHALTEN GEWESENE $STOFFE AUF . +	ANGEWANDTE	(1) employ	✓	T$(IN)
JSR12805	+ $BEI EINEM TOTALEN $GEWICHT VON 15 BIS 25 KG WURDEN $KALIBER BIS ZU 11,5 CM . +	ANGEWENDET	(1)		"
JSR07101	+ $ES WIRD DAHER UEBERCHLORSAURES $KALI , AUCH IN $BERUECKSICHTIGUNG SEINES HOEHEREN $PREISES WEGEN , FUER DEN $TREIBRAKETENBAU NUR SELTEN . + ***	ANGEWENDET	(1)		"
JSR06906	+ $FUER DIE $GUETE DIESER $KOHLE WIRD IN DER $FEUERWERKEREI FOLGENDE $FAUSTREGEL : +	ANGEWENDET	(2) apply	✓	T$(AB)
JSR12101	+ $AUCH HIER WURDE DAS $PRINZIP DER LOCKEREN $LADUNG , WOBEI ALS ACHSIALE $FUEHRUNG EIN DURCHLOECHERTES $ROHR ZUR $VERWENDUNG GELANGT . +	ANGEWENDET	(2)		"
JSR03203	+ $BEI GENAUER $BETRACHTUNG MUSS FUER DEN GANZEN $VORGANG JEDOCH DAS SEHR WICHTIGE $GESETZ << VON DER $ERHALTUNG DES $SCHWERPUNKTES BEI FREIEN $SYSTEMEN >>	ANGEWENDET	(2)		"
JSR10808	WERDEN . + + $ES IST DIES DIE FUER HOCHWERTIGE $STEIGRAKETEN AM MEISTEN	ANGEWENDETE	(1)		T$(IN)
JSR01314	$AUSFUEHRUNG . + + $WO ES AUS SE				

3.1.2 Coding of Rules

In a second pass through the concordance, linguists examined the set of occurrences of each word, disregarding different inflectional endings. No difficulties arose where only one translation was necessary to account for all occurrences of the word in the text, and where no other obvious translations were possible in the general usage of the word. For such pairs of German-English translation equivalents four separate rules were written:

a. A German dictionary rule for use in the automatic analysis of the German text. Such rules contain the features of the classified item and its selection restrictions in subscript format, as shown in Fig. III.2. The definition of the features is given under 3.2 below. 5,050 rules of this type were coded. Each rule represents all readings of a particular word stem. As can be seen in the example, all prefix-verb combinations for a particular verb are combined in one rule.

[For Fig. III.2 refer to following page]

b. A German normal form (or transfer) rule, which associates a normal form name (meaning) with the dictionary rule for the classified item. For convenience' sake, the normal form name is usually the canonical form of the classified item in English. Additional subscripts identify the German item covered by the rule and its category symbol, again as an aid to the human reader. Examples of German normal form rules are shown in Figure III.3. 5,150 such rules were coded.

[For Fig. III.3 refer to page III.5]

c. An English normal form rule whose name is identical to that of the German normal form rule, but which refers to an English (rather than a German) dictionary rule by the number on its right side. 4,150 such rules were coded. Examples of English normal form rules are shown in Fig. III.4.

[For Fig. III.4 refer to page III.6]

d. An English dictionary rule, whose identifying number is the one shown on the right side of the English normal form rule. Like the German dictionary rule, it assigns features and selection restrictions to the classified English item being rewritten. These rules were employed in the automatic generation of the output sentence. Examples of English dictionary rules are shown in Fig. III.5.

[For Fig. III.5 refer to page III.7]

Fig. III.2

RECOGNITION DICTIONARY


```
C 4383      V V                    • SEGL
D 1         • CL(1)
T GER DICT  • PX(LA,DAMIN)
            • FS(A)/
            • TS(AL)/
            • FO(LA)/
            • TO(LA)/
            • DA(CIR)/
            • AX(S)/
            • IO(LA)
            • IS(S)
            T 1,2

C 4384      V V                    • SEM
D 1         • CL(31)
T GER DICT  • PX(AR#VCR#VOR#VOR#
            VOR#US#AN#AUS#LA#L
            A#LA)/
            • FS(A#A#A#N#N#N#N#N
            #N#A)/
            • TS(HU#HL,AR#HU,AB#
            HU#HU#HL#AR#HU#HU#
            HU)/
            • FO(VCA#A#A,ALS#A,F
            UER#CL#A,ALS#LA#A,
            CL#A,IN2,CL#AUF1,C
            L)/
            • TO(AR#HL,AR#IN,IN#
            F,E#TH#IN,IN#LA#E,
            TH#AR,HI#TH)/
            • RA(LA#LA#LA#LA#LA#
            LA#HAN#LA#LA#LA)/
            • IO(C#O#C,02#G,02#O
            #O,C2#L#,O#O,02#O)
            • IS(S)
            T 1,2
            T 1,4
            T 1,5
            T 1,6

C 4564      V V                    • SESS
D 4         • CL(2)
T GER DICT  • PX(LA)/
            • FS(A)/
            • TS(HU)/
            • FO(LA)/
            • TO(LA)/
            • IO(LA)
            • IS(S)
            T 1,2
```

Fig. III.3

PRODJCTION TRANSFER ZERO

```
C 3448      V CONNECTION          C 22779
T GER TRAN  $ 2.)                 $ 10
            A CAT('N)
            N TM(ZUSAMMENHANG)

C 1414      V CONSCIOUS           C 10428
T GER TRAN  $ 2.)                 $ 10
            A CAT(A)
            N TM(AUFMERKSAM)

C 4653      V CONSEQUENCE         C 20426
T GER TRAN  A CAT(N)
            N TM(FOLGE)

C 3087      V CONSEQUENTLY        C 6041
T GER TRAN  A CAT(AV)
            N TM(DEMNACH)

C 3105      V CONSEQUENTLY        C 6042
T GER TRAN  A CAT(AV)
            N TM(DEMZUFOLGE)

C 3588      V CONSEQUENTLY        C 6120
T GER TRAN  A CAT(AV)
            N TM(INFOLGEDESSEN)

C 44        V CONSIDER            C 4384
T GER TRAN  . FO(1.0)             $ PX(AN)
            A CAT(V)              $ FO(ALS)
            N TM(ANSEHEN)

C 685       V CONSIDER            C 4384
T GER TRAN  . FO(0)               $ PX(VOR)
            A CAT(V)              $ TS(HU)
            N TM(VORSEHEN)        $ FO(A.*X)

C 2817      V CONSIDER            C 5020    C 20164   C 54      C 149     C 9000.65  C 4203
T GER TRAN  . FO(0)                         A 1       A 1       A 1       A 1        A 1
            A CAT(V.P)                                R 3       B 4       B 5
            N TM(INS.AUGE.FASSEN
              )

C 1404      V CONSIDERABLE        C 10417
T GER TRAN  A CAT(A)
            N TM(ANSEHNLICH)

C 1445      V CONSIDERABLE        C 10460
T GER TRAN  A CAT(A)
            N TM(BETRACHTLICH)

C 1516      V CONSIDERABLE        C 10538
T GER TRAN  A CAT(A)
            N TM(ERHEBLICH)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.4

PRODUCTION TRANSFER ZERO


```
C 1390      V CONSCIOUS          C 10442
T ENG TRAN  $ P=1                $ IO
            A CAT(A)

C 6058      V CONSEQUENCE        C 23069
T ENG TRAN  A CAT(N)

C 2911      V CONSEQUENTLY       C 6038
T ENG TRAN  A CAT(AV)

C 14        V CONSIDER           C 4432
T ENG TRAN  * FO(O,FOR)          $ FO(O,FOR
            A CAT(V)             )

C 18        V CONSIDER           C 4432
T ENG TRAN  * FO(O,AS)           $ FO(O,AS)
            A CAT(V)

C 30        V CONSIDER           C 4432
T ENG TRAN  * FO(O)              $ FO(O,*X)
            A CAT(V)

C 42        V CONSIDER           C 4432
T ENG TRAN  * FO(O,O)            $ FO(O,O)
            A CAT(V)

C 682       V CONSIDER           C 4432
T ENG TRAN  * FO(O,AJ)           $ FO(O,AJ)
            A CAT(V)

C 1380      V CONSIDERABLE       C 10432
T ENG TRAN  A CAT(A)

C 2150      V CONSIDERABLY       C 6063
T ENG TRAN  A CAT(AV)

C 2536      V CONSIDERATION      C 21873
T ENG TRAN  $ P=1                $ IO
            A CAT(N)

C 165       V CONSIST            C 4164
T ENG TRAN  A CAT(V)

C 2097      V CONSISTENTLY       C 6053
T ENG TRAN  A CAT(AV)

C 1600      V CONSPICUOUS        C 10661
T ENG TRAN  A CAT(A)

C 895       V CONSTANT           C 10616
T ENG TRAN  A CAT(A)

C 195       V CONSTITUTE         C 4194
T ENG TRAN  A CAT(V)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.5

RECOGNITION DICTIONARY


```
C 4260      V V                          • CONDUCT
T ENG DICT  • CL(3)
            • ON(C)
            • PR(LA)
            • FS(N)
            • TS(HU)
            • FO(0.TO=ARDS≠0≠0)/
            • TO(IN.IN≠IN≠AB)/
            • RA(LA≠DIR≠LA)/
            • IO(0.02≠0≠0)
            • IS(S)
            T 1.6

C 4762      V V                          • CONNECT
T ENG DICT  • CL(3)
            • ON(C)
            • PR(LA)
            • FS(N)
            • TS(HU)
            • FO(0.WITH.THROUGH.
              LA≠0)/
            • TO(IN.IN.IN.LA≠IN)
              /
            • IO(0.02.03.LA≠0)
            • IS(S)
            T 1.6

C 4432      V V                          • CONSIDER
T ENG DICT  • CL(3)
            • ON(C)
            • PR(LA)
            • FS(N)
            • TS(HU)
            • FO(0.AJ.0≠0.AJ.0≠0
              .FOR.LA≠0.AS≠0.AS)
              /
            • TO(AB.AR≠IN.IN≠IN.
              AB.IN.AB.LA≠AB.AB≠
              IN.IN)/
            • IO(0.02≠0.02≠0.02.
              LA≠0.02≠0.02)
            • IS(S)
            T 1.6
```

III.7

92

UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

4,400 English dictionary rules were coded. As in German, each rule represents all readings of a word stem which occur in the text. For adjectives and nouns, additional obvious readings and translations not occurring in the text were also included. Verb rules contain only the features necessary for analysis and translation of the text since these entries were complex enough to test the feature system and the algorithms. Verb-adprep combinations of a particular verb are combined in one rule. Only the main stems of irregular verbs (i.e., the stem of the canonical form or lemma) were coded. The rules for stem variants, including stems with an elided letter E (as in *handel/n - handl/e*), were subsequently generated automatically from a list of irregular verb stems. The features assigned by the lexicographer to the main stem were automatically assigned to its stem variants, with the exception of inflectional paradigmatic information, which was supplied by the lexicographers. Stem variants are recognized as allomorphs of the same morpheme by virtue of the fact that they have identical rule numbers (C-numbers) with different D-numbers ("duplicate numbers").

German verbs with inseparable prefixes and irregular verb stems were coded as derivational rules for the sake of economy. Thus, the rule for the German verb *entreissen* has the following format:

```
C 4912      V V          →      V ENT       C 4307
            $ 3.1                           $ CL
            + NP                            B
            + FS(N)
            + TS(AN)
            + IS(S)
            + FO(A.D)
            + TO(PO.AN)
            + IO(O.O2)
            + TM(ENTREISS
              EN)
            U 3
```

This rule concatenates the inseparable prefix *ent-* with the morpheme /reissen/, which is identified by the rule number C 4307. Since /reissen/ has two stem variants, *reiss* and *riss*, this one rule takes the place of two prefix-stem combinations; in other cases, one rule replaces three or four, representing a considerable savings in the dictionary.

3.1.3 Rule Coding for Ambiguous Items

In all cases where more than one English translation equivalent was needed for the occurrences of a German word in the text (or for the translation of other, common readings of that word), the linguist determined what criteria enable the human

translator to choose one of the several possible translation equivalents. If the criteria were of a linguistic nature (usually selection restrictions referring to other phrase or sentence constituents), those restrictions were coded as features of the classified item in the dictionary rules. These distinguishing features were then explicitly mentioned in the normal form rules, thus enabling the system to apply automatically the same criteria for the selection of a translation equivalent which are used by the human translator. For an example of such a case, refer to Fig. III.1, where the human translator indicated two translation possibilities for the German verb *anwenden* (here occurring in its forms *angewandte, angewendet* and *angewendete**): the English verbs *employ* and *apply*. The linguist established that *employ* is the appropriate translation for *anwenden* whenever that German verb occurs with a noun denoting an inanimate physical object in its object slot, while *apply* is chosen when the German verb occurs with an abstract object. Thus we get

employ coal
employ calibers (of guns)
employ potassium perchlorate
employ a specific model

and *apply a rule of thumb*
apply a principle

because the German nouns for *coal, caliber, potassium perchlorate,* and *model* were given the feature "inanimate physical object" and the nouns *rule of thumb* and *principle* the feature "abstract, countable". The rules below show how such disambiguating information is indicated in dictionary and normal form rules.

[For Figs. III.6.1/9 to the following pages]

Cf. Section II for a description of the algorithm which actually makes use of the features coded in the dictionaries.

However, in all cases where the human translator uses extralinguistic information in his choice of translation equivalents, the linguist was unable to indicate this information in the dictionary and normal form rules. The system therefore could not make a choice, and selected all possible translations given. An

*This particular verb did not occur in discontinuous structures, as e.g., *wendeten... an.* A great number of other prefixed verbs did occur discontinuously; their different forms were considered for classification and translation simultaneously.

Fig. III.6.1

```
RECOGNITION DICTIONARY                                          PAGE NO.  191

          C 4576      V V                          • WEICH
          T GER DICT  • CL(49)
                      • PX(LA)
                      • FS(N)
                      • TS(E)
                      • FO(O)
                      • TO(E)/
                      • IO(O)
                      • IS(S)

          C 4531      V V                          • WEIS
          D 1
          T GER DICT  • CL(50)
                      • PX(NACH#HIN#LA#AUF
                        #AUF)/
                      • FS(N#N#A#N#N)/
                      • TS(HU#HL#HU#PO#NT)
                        /
                      • FO(A#AUF)#A#A#A)/
                      • TO(AR#AR#AR#AB,IN#
                        AR)/
                      • RA(T*DU#LA#LA#LA#L
                        A)/
                      • TY(LA#LA#LA#NP#NP)
                        /
                      • IO(C#C#C#O#O)
                      • IS(S)
                      T 1.2
                      T 1.5

--->      C 4532      V V                          • WEND
          D 1
          T GER DICT  • CL(60)
                      • PX(AN#A#A#AUF#ZU)/
                      • FS(N#N#A#N)/
                      • TS(HU#HL#HU#HU)/
                      • FO(A.ZU#A.FUER.LA#
                        A#A.O)/
                      • TO(IN.AR#IN.AR.IN.
                        AR.LA#IN.AR#R.E)/
                      • TY(LA#LA#LA#NP)/
                      • IO(C.02#O.02*LA#O#
                        LA.C)
                      • IS(S)
                      T 1.2
                      T 1.5
                      T 1.6
```

Fig. III.6.2


```
C 242          V EMPFD                C 4077
T GER TRAN     A CAT(V)               $ PX(EIN)
               N TM(EINRETTEN)

C 560          V EMERGE               C 4223
T GER TRAN     • FU(FROM)             $ PX(HERVO
               • TO(HU•CO)              R)
               A CAT(V)               $ TS(HU•CO
               N TM(HERVORGEHEN)        )
                                      $ FO(AUS)

C 4059         V EMISSION             C 20530
T GER TRAN     A CAT(N)
               N TM(AUSSTRAHLUNG)

C 4663         V EMPHASIZE            C 5036    C 22576   C 54      C 149      C 9000,69  C 4403
T GER TRAN     • FU(O)                          A 1       A 1       A 1        A 1        A 1
               A CAT(V•P)                                 B 3       B 4        B 5
               N TM(IN•DEN•VORDERGR
                 UND•STELLEN)

C 1511         V EMPIRICAL            C 10533
T GER TRAN     A CAT(A)
               N TM(ERFAHRUNGSGEMAE
                 SS)

C 161          V EMPLOY               C 4054
T GER TRAN     • FU(O•FOR•LA)         $ TO(HU)
               $ 2•3                  $ IO
               A CAT(V)
               N TM(BESCHAEFTIGEN)

→ C 48         V EMPLOY               C 4532
T GER TRAN     • FU(O•FOR•LA)         $ PX(AN)
               $ 2•3                  $ TO(IN)
               A CAT(V)               $ IO
               N TM(ANWENDEN)

C 4869         V EMPLOY               C 5042    C 21179   C 54      C 149      C 9000,68  C 4247
T GER TRAN     • FU(O•FOR•LA)         $ IO                A 1       A 1        A 1        A 1
               $ 2•3                                      B 3       B 4        B 5
               A CAT(V•P)
               N TM(ZUM•EINSATZ•KOM
                 MEN)

C 3633         V EMPLOYEE             C 20069
T GER TRAN     A CAT(N)
               N TM(ANGESTELLTER)

C 1021         V EMPTY                C 10035
T GER TRAN     A CAT(A)
               N TM(LEER)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.6.3


```
C 4205      V APPLICABILITY          C 20714
T GER TRAN  A CAT(N)
            N TM(VERWENDUNGSZWEC
              K)

C 3210      V APPLICABILITY          C 22550
T GER TRAN  $ 2.1                    $ IO
            A CAT(N)
            N TM(VERWENDBARKEIT)

C 3661      V APPLICATION            C 20099
T GER TRAN  $ 2.1                    $ IO
            A CAT(N)
            N TM(ANWENDUNG)

C 2294      V APPLICATION            C 21620
T GER TRAN  A CAT(N)
            N TM(ANMELDUNG)

C 197       V APPLY                  C 4089      C 4560      C 9200
T GER TRAN  $ 2.1                    $ IO                    A 1
            A CAT(V)                 $ TO(*R)
            N TM(REZIFHEN)

C 49        V APPLY                  C 4532
T GER TRAN  $ 2.3                    $ PX(AN)
            A CAT(V)                 $ TO(AB.AB
            N TM(ANWENDEN)             .LA)
                                     $ IO
                                     $ FO(A.FUE
                                       R.LA)

C 1582      V APPLY                  C 5013      C 20099     C 34        C 4206
T GER TRAN  $ 2.1                    $ IO(LA)                A 1         A 1
            A CAT(V.P)                                                   R 3
            N TM(ANWENDUNG.FINDE
              N)

C 546       V APPROACH               C 4322
T GER TRAN  . TS(HU)                 $ PX(HERAN
            A CAT(V)                   )
            N TM(HERANPRESCHEN)

C 3631      V APPROACH               C 20067
T GER TRAN  A CAT(N)
            N TM(ANFLIEGUNG)

C 1508      V APPROPRIATE            C 10530
T GER TRAN  $ 2.1                    $ IO
            A CAT(A)
            N TM(ENTSPRECHEND)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.6.4

PRODUCTION TRANSFER ZERO


```
C 235        V EMBED                C 4230
T ENG TRAN   A CAT(V)

C 655        V EMERGE               C 4560
T ENG TRAN   * FO(FROM)             $ FO(FROM)
             * TO(HU*CO)            $ TO(HU*CO
             A CAT(V)                 )

C 4055       V EMISSION             C 20524
T ENG TRAN   A CAT(N)

C 492        V EMPHASIZE            C 4481
T ENG TRAN   * FO(O)                $ FO(O)
             A CAT(V)

C 1486       V EMPIRICAL            C 10541
T ENG TRAN   A CAT(A)

C 46         V EMPLOY               C 4153
T ENG TRAN   * FO(O,FOR,LA)         $ FO(O,FOR
             $ P=2                    ,LA)
             A CAT(V)               $ TO

C 3616       V EMPLOYEE             C 20067
T ENG TRAN   A CAT(N)

C 1013       V EMPTY                C 10035
T ENG TRAN   A CAT(A)

C 4037       V ENCIRCLEMENT         C 20504
T ENG TRAN   A CAT(N)

C 245        V ENCLOSE              C 4240
T ENG TRAN   A CAT(V)

C 122        V ENCOUNTER            C 4121
T ENG TRAN   A CAT(V)

C 407        V ENCOURAGE            C 4398
T ENG TRAN   $ P=1                  $ IO
             A CAT(V)

C 3839       V ENCUMBRANCE          C 20261
T ENG TRAN   A CAT(N)

C 259        V END                  C 4254
T ENG TRAN   A CAT(V)               $ PX(LA)

C 4262       V END                  C 20532
T ENG TRAN   A CAT(N)

C 403        V ENDANGER             C 4396
T ENG TRAN   A CAT(V)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.6.5

PRODUCTION TRANSFER ZERO


```
C 927       V APPEAR            C 4318
T ENG TRAN  * FO(AS)            $ FO(AS)
            A CAT(V)

C 434       V APPEAR            C 4318
T ENG TRAN  * FO(LA)            $ FO(LA)
            A CAT(V)

C 1983      V APPEARANCE        C 23155
T ENG TRAN  A CAT(N)

C 3657      V APPLICABILITY     C 20100
T ENG TRAN  $ P=1               $ IO
            A CAT(N)

C 3648      V APPLICATION       C 20101
T ENG TRAN  $ P=1               $ IO
            A CAT(N)

C 149       V APPLY             C 4189
T ENG TRAN  $ P=1               $ IO
            A CAT(V)

C 561       V APPROACH          C 4529
T ENG TRAN  * TS(MU)            $ TS(MU)
            A CAT(V)

C 3514      V APPROACH          C 20065
T ENG TRAN  A CAT(N)

C 3203      V APPROACH          C 23041
T ENG TRAN  $ P=1               $ IO
            A CAT(N)
            N IM(PLIGHT)

C 1983      V APPROPRIATE       C 10538
T ENG TRAN  $ P=1               $ IO
            A CAT(A)

C 1376      V APPROXIMATE       C 10428
T ENG TRAN  A CAT(A)

C 133       V APPROXIMATELY     C 6295
T ENG TRAN  A CAT(AV)

C 1378      V APPROXIMATING     C 10793
T ENG TRAN  $ P=1               $ IO
            A CAT(A)

C 4500      V APRIL             C 22910
T ENG TRAN  A CAT(N)

C 1312      V AQUEOUS           C 10362
T ENG TRAN  A CAT(A)
```

UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.6.6


```
C 4153      V V                    • EMPLOY
T ENG DICT  • CL(3)
            • ON(V)
            • PA(LA)
            • FS(N)
            • TS(MJ)
            • FO(O#O,FOR,LA)/
            • TO(MJ#IN,AB#IN,LA)
              /
            • IO(O#O,O2#LA)
            • IS(S)
            T 1.6

C 4240      V V                    • ENCLOS
T ENG DICT  • CL(1)
            • ON(V)
            • PA(LA)/
            • FS(N)/
            • TS(IN)/
            • FO(O)/
            • TO(IN)/
            • IO(O)
            • IS(S)

C 4121      V V                    • ENCOUNTE
T ENG DICT  • CL(3)                  R
            • ON(V)
            • PA(LA)/
            • FS(N)/
            • TS(MJ)/
            • FO(O)/
            • TO(AB)/
            • IO(O)
            • IS(S)

C 4398      V V                    • ENCOURAG
T ENG DICT  • CL(1)
            • ON(V)
            • PA(LA)
            • FS(N#CL#N)/
            • TS(MJ#TH#AB)/
            • FO(O#O,LA)/
            • TO(AB#MJ,LA)/
            • IO(O#O,LA)
            • IS(S)
            T 1.7
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.6.7


```
C 4189      V V                 • APPLY
D 1         • CL(31)
T ENG DICT  • ON(V)
            • PX(LA)/
            • FS(N)/
            • TS(HU)/
            • FO(O.TO.LA)/
            • TO(AB.AB.LA)/
            • IO(O.O2.LA)
            • IS(S)
            T 1.7
            L

C 4529      V V                 • APPROACH
T ENG DICT  • CL(2)
            • ON(V)
            • PX(LA)/
            • FS(N)/
            • TS(HU)/
            • FO(LA)/
            • TO(LA)/
            • IO(LA)
            • IS(S)
            T 1.5

C 4056      V V                 • ARGU
T ENG DICT  • CL(1)
            • ON(V)
            • PX(LA)/
            • FS(N)/
            • TS(HU)/
            • FO(CL.LA)/
            • TO(TH.LA)/
            • IO(O.LA)
            • IS(S)
            T 1.7

C 4290      V V                 • ARIS
D 1         • CL(21)
T ENG DICT  • ON(V)
            • PX(LA)/
            • FS(N)/
            • TS(AB)/
            • FO(LA)/
            • TO(LA)/
            • IO(LA)
            • IS(S)
            T 1.5
            L
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.6.8

PRODUCTION DICTIONARY


```
        C 21158     V N                 • KOENIG
        T GER DICT  • CL(17)
                    • GD(M)
                    • TY(NU)

        C 21023     V N                 • KOEPF
        D 2         • CL(7)
        T GER DICT  • GD(M)
                    • TY(IN)

        C 21856     V N                 • KOERB
        D 2         • CL(7)
        T GER DICT  • GD(M)
                    • TY(IN)

        C 21160     V N                 • KOERN
        D 2         • CL(8)
        T GER DICT  • GD(N)
                    • TY(IN)

        C 21726     V N                 • KOERNCHE
        T GER DICT  • CL(13)              N
                    • GD(N)
                    • TY(IN)

        C 21161     V N                 • KOERNUNG
        T GER DICT  • CL(1)
                    • GD(F)
                    • TY(AB)
                    • NU(S)

        C 20855     V N                 • KOERPER
        T GER DICT  • CL(19)
                    • GD(M)
                    • TY(IN)

  →     C 21163     V N                 • KOHLE
        T GER DICT  • CL(36)
                    • GD(F)
                    • TY(IN)

        C 21164     V N                 • KOHLENGE
        T GER DICT  • CL(17)              BIET
                    • GD(N)
                    • TY(IN)

        C 21165     V N                 • KOHLENLA
        T GER DICT  • CL(19)              GER
                    • GD(N)
                    • TY(IN)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.6.9

PRODUCTION DICTIONARY

C 21688	V N	• PRESSE
T GER DICT	• CL(34)	
	• GU(F)	
	• TY(MO•CO,AB)	
C 21691	V N	• PRESSLUF
T GER DICT	• CL(1)	T
	• GU(F)	
	• TY(I'M•MS)	
C 21692	V N	• PRESSMET
T GER DICT	• CL(34)	HODE
	• GU(F)	
	• TY(AB)	
C 21693	V N	• PRESSNIT
T GER DICT	• CL(1)	I
	• GU(N)	
	• TY(IM•NM)	
C 21694	V N	• PRESSVER
T GER DICT	• CL(13)	FAHREN
	• GU(N)	
	• TY(AB)	
C 21695	V N	• PRESSVOR
T GER DICT	• CL(20)	RICHTUNG
	• GU(F)	
	• TY(IN)	
→ C 21696	V N	• PRINZIP
T GER DICT	• CL(32)	
	• GU(N)	
	• TY(AB)	
C 21697	V N	• PRITSCHE
T GER DICT	• CL(34)	
	• GU(F)	
	• TY(IN)	
C 21714	V N	• PROBE
T GER DICT	• CL(34)	
	• GU(F)	
	• TY(AB)	
C 23138	V N	• PROBESCH
T GER DICT	• CL(2)	IESSEN
	• GU(N)	
	• TY(AB)	

UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

example for this problem is the German adjective *gering*, which is translated as *low*, *small*, or *little*, depending on the type of noun it modifies:

geringe Kosten = *low cost*
geringe Mengen = *small amounts*
geringe Energie = *little energy*

No distinction can be made because all three nouns which occur in the examples above fall into the same semantic class in our present classification system: "abstract". It is possible that a finer but still practical classification system can be worked out which would enable the linguist to select the correct translation from the three shown here. To facilitate further research into this problem, three lists were made of ambiguous German nouns, verbs, and adjectives, respectively, which we were not able to disambiguate.

In addition, separate lists were made of lexical items representing different types of problems:

a. lists of German noun compounds, divided into two sub-lists of compounds which can be automatically analyzed or generated from their components and those which cannot;

b. a list of adjective compounds and adjective phrases, e.g., *blassblau - pale blue;*

c. lists of German and English verbal lexical collocations, divided into a list of collocations which have internal variable slots (e.g., *etwas einer* [adjective slot] *Pruefung unterziehen - test something* [adverb slot]) and those which do not have such a variable slot (e.g., *jemandem etwas vor Augen fuehren - demonstrate something to somebody; zutreffen - be the case*);

d. a list of verbs which must be translated in a different voice, e.g., *gelten als* (active voice) - *be considered* (passive voice) or *sich stuetzen auf - be based on;*

e. lists of lexical items which must be translated into English items of another category, subdivided into a list of German nominalized adjectives (e.g., *der Angestellte - the employee*), German nouns to be translated as English nominalized verb forms (e.g., *der Abschuss - the launching*), and German adjectives to be translated as English nouns (e.g., *feindlich - enemy (lines)*);

f. lists of German and English multi-word prepositions, e.g., *nach X hin - toward X; in Bezug auf - with respect to;*

g. lists of German and English multi-word adverbs, e.g., *nach wie vor - still; im Durchschnitt - on the average.*

Most of the problems represented by these lists were dealt with after the bulk of the lexicographic work (as described above) was completed. The solutions arrived at are described below under pt. 3.3.

Finally, one other list of a different type might be worth mentioning, namely the list of ill-formed phrases and sentences which occur in the text. This list, also compiled by the lexicographers during their second pass through the concordance, is surprisingly long. It was subdivided into lists of incorrect punctuations, misspellings, incorrect inflections, incorrect word choice, and incorrect syntax. These ill-formed sentences could not be automatically analyzed or translated with our present grammars.

Idiosyncratic usage of German by the author of our text, such as *fusses in* (rather than *auf*), was accounted for in the dictionary rules; thus, *fussen* was given the feature FO(IN2,IN+CL) for "takes an obligatory object complement which consists of the preposition *in* followed by a dative noun phrase (IN2) or by a clause (IN+CL)", as in *"Diese Schwierigkeit fusst darin, dass . . ."* (The particular type of clause allowed in this slot is specified under the subscript TO). Thus, analysis and translation of the author's idiosyncratic use of German was possible.

(It will be interesting to investigate whether such incorrect selection of the preposition of prepositional complements occurs in other texts. If so, we might establish a rule which permits a writer to select from "similar" prepositions when in doubt. Here, it seems, the author of the text selected the locative *in* instead of the locative *auf*.)

3.2 Lexical Features

The following is a description of the system of syntactic and semo-syntactic features used in the classification of the word stems which occur in the text *Raketenantriebe* and their English translation equivalents. It is essentially the same set of features which has been used by the Center's lexicographers in their lexicographic work since 1971. Earlier descriptions can be found in RADC-TR-72-47 Technical Report 1972 and RADC-TR-73-309 Final Technical Report October 1973. Entries from the LRC dictionaries are shown at the end of each section to illustrate the use of the classification described.

3.2.1 Noun Classification

Nouns are classified mainly with respect to their semantic type. This information is coded in form of the subscript TY("type") with one of the following values:

Human (HU)
Animal (AL)
Plant (PL)
Inanimate physical object (IN)
Abstract (AB)

An additional class NT was introduced for those nouns which posed problems as to their membership in the classes AB or IN; examples are the nouns *current*, *wind*, and others. The nouns in this class denote mainly physical phenomena which can be perceived by the senses and therefore function as object complements for such verbs as *hear*, *measure*, *register*, but which do not function as object complements for verbs like *put*, *paint*, etc., which require physical-object-type complements.

Several of these noun classes comprise subclasses. Thus, a subclass of HU is "Title" (T), for nouns like *Mr.*, *Professor*, etc. Nouns in this class help to identify the following word as a human proper name.

The noun class IN (Inanimate Physical Object) contains two subclasses: "Body Part" (BP) and "Machine" (MA). The class BP was established to account for the fact that the nouns in this class, such as *leg*, *hair*, etc., are, in some contexts, preceded by possessive pronouns in English, while some other languages, among them German, use the definite article in the same context:

Er oeffnete die Augen - He opened his eyes

The subclass MA contains nouns like *jet*, *rocket*, *computer*, etc., which may function as subjects of such verbs as *travel* or *calculate*, i.e., verbs which normally require human subjects.

The class AB (Abstract) has three subclasses. The first of these is "Count" (CN), containing abstract nouns which, as distinct from other abstract nouns, may not occur in the singular without a determiner. Examples are *discovery*, *idea*, *prognosis*. Another subclass of Abstract is "Unit" (UN). It contains the limited number of nouns which, if preceded by a quantifier, may serve as adverbials, e.g., *year*, *kg*, cm^3, but also *man*:

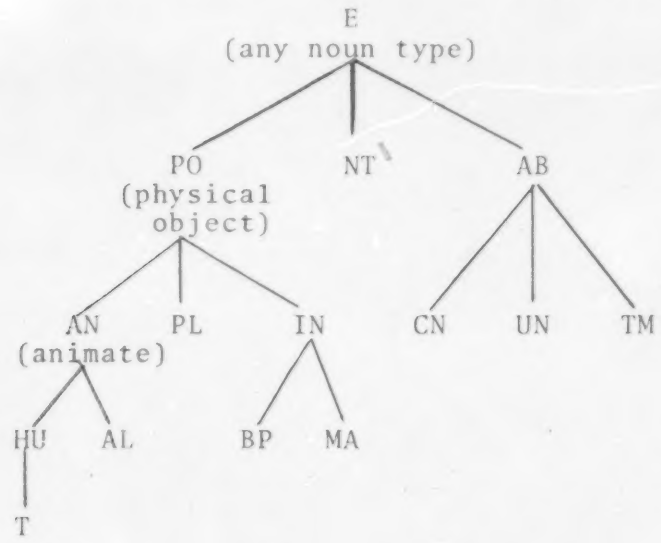
It is a few years old.

It weighs 5 kg.

The team is eleven men strong.

The last subclass of Abstract is "Time" (TM), which marks nouns like *day*, *year*, etc. This noun class serves to disambiguate some prepositions. For example, the preposition *in* forms a time adverbial if used with a "time noun" (*in the year 1935*) and a locative adverbial if used with a noun denoting an inanimate object, non-mass (*in this pipe*).

This noun classification may be represented in the following tree, to which the system has automatic access:



Thus, if a noun like *author* is clasified as "human", it is automatically recognized by the system as a member of the increasingly larger classes "Animate" and "Physical Object". We are considering the possibility of modifying the tree by placing PL (Plant) under the node IN (Inanimate Physical Object) because the class PL seems to be included in the selection restrictions of most verbs and adjectives which specify the class IN. This modification of the tree would simplify coding of such selection restrictions; the symbol IN would thus replace the earlier IN, PL.

Three additional noun classes are not shown in this tree because they may be combined with any of the basic noun classes. They are "Name" (NM), "Collective" (CO), and "Mass" (MS). "Collective" refers to nouns which are plural in meaning but singular for grammatical purposes, and "Mass" refers to inanimate nouns, usually materials or substances, which may be used without a determiner in the singular. Each of these classes may be combined with any of the basic classes (by means of a "+"), resulting in a new semo-syntactic subclass of that basic noun class. Examples are:

acetylene - TY(IN+MS)
government - TY(HU+CO)
John - TY(HU+NM)

In addition to one of the mentioned classes, the feature "Quantifier" (QU) is assigned to all those nouns which frequently function as the syntactic head of a noun phrase, followed by an

of-phrase (in German by a genitive or *von*-phrase) whose noun is the semantic head of the whole noun phrase construction. Examples of such nouns which are also members of the class "Abstract" are *third, series, number;* nouns which denote inanimate physical objects and are members of the class QU are *glass, container, box.* In German, this latter group of nouns also permits modifications by *mit-, voll-,* and *voller*-phrases, e.g., *eine Schachtel voll(er) Pulver, ein Glas mit Wasser.*

Ambiguous nouns, such as the English *conductor*, are given the features of all their readings, in this case TY(HU,IN). If the context within the sentence allows, the item is disambiguated in Syntactic Choice (cf. Section 2.2.3.3 on page II).

German nouns were given gender information unless gender is irrelevant, as in the case of nouns which exist only in their plural forms, such as *Ferien, Kosten.* These nouns were marked as plurale tantum nouns by the subscript TT(P).

English nouns were also subclassified as to onset to guarantee the correct choice of the proper indefinite article in the output sentence.

The valences of those nouns which take complements were specified in the noun entries as to their required syntactic form and semantic type. Under the subscript FC ("Form of Complement"), a specific preposition was coded if the complement is a prepositional phrase; the symbol CL was coded if the complement is a clause. The specific values associated with the subscript TC ("Type of Complement") indicate the semantic type of the complement. They refer to the head noun of the prepositional phrase or the particular type of the complement clause. (The description of the different types of complement clauses in our feature system is given on p. III.27, under 3.3.3, Verb Classification.)

Finally, a value specifying the inflectional paradigm was coded in the dictionary rule for each German and English noun stem. It may be added here that the experiment, as originally planned, was meant to produce translation output in so-called English standard strings, sentences in which the word order is standardized and each word is represented by its canonical form. (A special English grammar, the rearrangement grammar, later permutes the words into surface order and selects the correctly inflected word form or stem variant of each canonical form.) For this reason, English nouns, verbs, and adjectives were originally coded without this inflectional paradigmatic information. Later, inflectional information was added to all English noun, verb, and adjective rules to make the output as close to surface English as possible.

As an illustration, some German and English dictionary rules for nouns may follow. [For Figs. III.7.1/2 refer to following pages]

Fig. III.7.1


```
C 23143     V N                 • ATTEMPT
T ENG DICT  • CL(1)             P
            • ON(V)
            • TY(CN)
            • FC(LA=CL)
            • TC(LA=MI)
            • IG(LA=0)
            • LC(NP)

C 20153     V N                 • ATTENTIO
T ENG DICT  • CL(4)               N
            • ON(V)             P
            • TY(CN)
            • LC(N)

C 22865     V N                 • AUGUST
T ENG DICT  • CL(4)             P
            • ON(V)
            • TY(AB=MO=NM=M)
            T 1=3

C 21556     V N                 • AUSTRIA
T ENG DICT  • CL(4)             P
            • ON(V)
            • TY(IN=NX)

C 21557     V N                 • AUSTRIAN
T ENG DICT  • CL(1)             P
            • ON(V)
            • TY(MO=M,F)

C 21987     V N                 • AUTHOR
T ENG DICT  • CL(1)             P
            • ON(V)
            • TY(MO=M,F)

C 21025     V N                 • AUTHORIT
U 2         • CL(10)              I
T ENG DICT  • ON(V)             P
            • TY(MO=CO)
            L

C 21025     V N                 • AUTHORIT
U 1         • CL(4)               Y
T ENG DICT  • ON(V)             P
            • TY(MO=CO)
            L

C 20213     V N                 • AUTOMOBI
T ENG DICT  • CL(3)               L
            • ON(V)             P
            • TY(MA)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.7.2

PRODUCTION DICTIONARY


```
        C 20012     V N                 • ARNEIGUN
        T GER DICT  • CL(20)              G
                    • GU(F)
                    • TY(AB)
                    • FC(LA,GFGEN)
                    • TC(LA,AR)
                    • IO(LA,O)
                    T 1..

        C 20013     V N                 • ABSCHEID
        T GER DICT  • CL(20)              UNG
                    • GN(F)
                    • TY(AB)
                    • LC(PP)

        C 20014     V N                 • ABSCHLUS
        T GER DICT  • CL(5)               S
                    • GU(M)
                    • TY(AB,IN)
                    T 1.3

        C 22902     V N                 • ABSCHLUS
        T GER DICT  • CL(19)              SDECKEL
                    • GU(M)
                    • TY(IN)

        C 22903     V N                 • ABSCHLUS
        T GER DICT  • CL(15)              SFLANSCH
                    • GU(M)
                    • TY(IN)

        C 22904     V N                 • ABSCHLUS
        T GER DICT  • CL(34)              SPLATTE
                    • GU(F)
                    • TY(IN)

        C 20015     V N                 • ABSCHNIT
        T GER DICT  • CL(17)              T
                    • GU(M)
                    • TY(IN,AR,NT)
                    • FC(LA,UEBER1)
                    • TC(LA,AR,IN)
                    • IO(LA,O)
                    T 1.4
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

3.2.2 Adjective Classification

The most important feature coded in adjective entries indicates the semantic noun type with which the adjective may co-occur. This information is coded under the subscript "Type of Modificand" (TM). Since the algorithm has access to the value tree of semantic noun classes shown above on p. III.22, these selection restrictions (and those of verbs) can refer to any node of that tree instead of having to list all noun classes with which a particular adjective may co-occur. Thus, those adjectives which can modify any type of noun, e.g., *good*, were coded with the feature TM(E), the top node of the value tree of semantic noun classes. Other adjectives are more restricted in use and therefore refer to lower nodes in the tree, as e.g., *liquid*, which has the feature TM(IN).

Those adjectives which may modify not only nouns but also clauses were coded with the subscript "Form of Modificand" (FM) and the associated value "Clause" (CL). In such dictionary rules the subscript "Type of Modificand" specifies the specific clause type and was coded with any of the following values, as the adjective being classified allows: "*that*-clause" (TH), "Marked Infinitive" (MI), "Interrogative Clause" (ICL), "*for-to* Construction" (FT), or "impersonal *it*" (IT). (Cf. adjective rules below for examples.)

As with nouns, those adjectives which take object complements were coded with subscripts to indicate the syntactic form and the semantic type of the complement. The subscript "Form of Object Complement" (FO) specifies in its value (or values) whether the complement is a clause, a prepositional phrase, or a noun phrase (the latter in German only). For noun phrase complements, the case governed by the adjective was coded (e.g., *bekannt* FO(D), for "dative"); for prepositional objects, the specific preposition. If the prepositional object consists of a preposition followed by a clause, the value "Clause" was combined with the specific preposition, e.g., *dependent* FO(ON,ON+CL), as in

X is dependent on Y.
X is dependent on the fact that . . .
X is dependent on how Y is done.

In English transfer, the constant *the fact* is inserted whenever a preposition is followed by a *that*-clause in English.

In all such rules, the values coded with the subscript "Type of Object Complement" (TO) indicate the semantic noun class which the head noun of phrasal complements must belong to, or the type of clausal complements.

The small number of adjectives which require an adverbial complement were coded with the subscript "Required Adverbial" (RA), whose values indicate which particular type of adverbial complement is required. The adverb classes available for this purpose are "Locative" (L), subclassified into "Stative" (ST) and "Directional" (DIR); "Time" (T), subclassified into "Punctual" (PU), "Duration" (DU), and "Frequency" (FR); "Manner" (M); and "State" (STATE). An example of such an adjective is the German *wohnhaft*, which requires a stative locative adverbial complement.

Another small class of adjectives was given the feature TY(MSR), "Measurable", to indicate that these adjectives, such as *wide, old,* etc. may be modified by an adverbial of the form quantifier + quantifier-noun, as in 5 *inches wide.* (Cf. also 3.2.1, Noun Classification.)

In general, present and past participles which function as adjectives in the text were only entered as verb stems. The morphological rules interpret these verb forms as adjectives. Those adjectival participles, however, whose features and meaning (and translation) differ from that of the underlying verb stem were coded as adjectives. They were given the additional subscript SP with the value "Past Participle" (PAPL) or "Present Participle" (PRPL).

All English adjectives, like English nouns and verbs, were given onset information. Both German and English adjectives were coded with a subscript-value combination which indicates the proper inflectional paradigm for each entry.

The following adjective rules, taken from our German and English dictionaries, may serve to illustrate the features mentioned.

[For Figs. III.8.1/2 refer to the following pages]

3.2.3 Verb Classification

Verbs were classified mainly in terms of their co-occurrence restrictions. As in the classification of adjectives, the restrictions for object complements are stated in the form of two subscripts with associated values, which indicate the "Form of the Object Complement" (FO) and the "Type of the Object Complement" (TO). As in the adjective description (cf. 3.2.2 above, for details), the subscript FO specifies the particular case governed by the verb, which may be nominative, genitive, dative, or accusative (in German only), or a preposition. In the latter instance, the specific preposition(s) is (are) coded as values of FO. For verbs which take clausal complements, the subscript FO has the value CL ("Clause"), which may be preceded by a preposition if the clause complement is to be introduced by a preposition (cf. the rule for *sehen* in Fig. III.2, which allows for constructions like

Fig. III.8.1

```
PRODUCTION DICTIONARY                                          PAGE NO.  2

        C 10512     V A                     * ACTUAL
        T ENG DICT  * CL(01)                P
                    * TM(IN*AH)
                    * ON(V)

        C 10529     V A                     * ACUTE
        T ENG DICT  * CL(01)                P
                    * TM(AB)
                    * ON(V)

        C 10395     V A                     * ADDITION
        T ENG DICT  * CL(29)                  AL
                    * TM(AB*IN)             P
                    * ON(V)

        C 10319     V A                     * ADJUSTAB
        T ENG DICT  * CL(05)                  L
                    * TM(AB*IN)             P
                    * ON(V)
                    ¬

        C 10356     V A                     * ADVANTAG
        T ENG DICT  * CL(01)                  EOUS
                    * TM(AB#MI)/            P
                    * FM(NO#CL)
                    * ON(V)

        C 10142     V A                     * ADVISABL
        T ENG DICT  * CL(05)                P
                    * TM(AB#MI*FT)/
                    * FM(NO#CL)
                    * ON(V)
                    ¬

        C 10576     V A                     * AERODYNA
        U 1         * CL(03)                  MIC
        T ENG DICT  * TM(AH)                P
                    * ON(V)

        C 10575     V A                     * AERODYNA
        U >         * CL(01)                  MICAL
        T ENG DICT  * TM(AB)                P
                    * ON(V)

        C 10416     V A                     * AGGRESSI
        T ENG DICT  * CL(01)                  VE
                    * TM(AB*AN)             P
                    * ON(V)

        C 10427     V A                     * AGREEABL
        T ENG DICT  * CL(05)                P
                    * TM(AB*MU)
                    * ON(V)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

Fig. III.8.2

```
PHODJCTION DICTIONARY

        C 10231     V A                 • TAEGLICH
        T GER DICT  • CL(A)
                    • TM(AB)

        C 10844     V A                 • TAETIG
        T GER DICT  • CL(5)
                    • TM(MU)
                    • FU(LA,ALS)
                    • TU(LA,AR,MU)
                    • IO(LA,O)
                    T 1.3

        C 10232     V A                 • TATKRAFF
        T GER DICT  • CL(5)               TIG
                    • TM(MU,AB)

        C 10233     V A                 • TATSAECH
        T GER DICT  • CL(A)               LICH
                    • TM(AB)

        C 10234     V A                 • TECHNISC
        T GER DICT  • CL(11)              H
                    • TM(AH)

        C 10758     V A                 • TECHN.
        T GER DICT  • CL(20)
                    • TM(AB)

        C 10235     V A                 • TEILWEIS
        T GER DICT  • CL(11)              E
                    • TM(AB)

        C 10110     V A                 • TELEGRAP
        T GER DICT  • CL(11)              HISCH
                    • TM(AH,NT)

        C 10236     V A                 • TEUER
        D 1         • CL(18)
        T GER DICT  • TM(IN,AB)

        C 10236     V A                 • TEUR
        D 2         • CL(7)
        T GER DICT  • TM(IN,AB)

        C 10237     V A                 • THEORETI
        T GER DICT  • CL(11)              SCH
                    • TM(AB)

        C 10238     V A                 • THERMISC
        T GER DICT  • CL(11)              H
                    • TM(AB,IN)
```


UNIVERSITY OF TEXAS AT AUSTIN COMPUTATION CENTER

darauf sehen, dass . . . in its last column). If the verb takes an adjective complement, the value AJ is coded. The subscript TO specifies the semantic noun class to which the noun complement must belong and the form which clausal complements may take. The types of clause complements which can be specified here are "*that*-clause" (TH); "Marked Infinitive" (MI); "Interrogative Clause" (ICL); "Interrogative Pronoun + Marked Infinitive" (IMI), "*for-to* Construction" (FT), "Gerund" (GR), "Accusative-Infinitive Construction" (ACI), the last four pertaining to English verbs only; "Past Participial Complement" (PAPL); and "Main Clause in the Subjunctive" (SJ, German only). In addition to noun and sentence types, the subscript TO may specify that the complement must be a reflexive (R) or reciprocal (RP) pronoun.

Generally, the same types of restrictions were specified for the subject of the verb under the two subscripts "Form of Subject" (FS) and "Type of Subject" (TS). Some descriptors, such as "Reflexive" and "Reciprocal", of course, never occur in the description of the subject. (The terms "subject" and "object" refer to the "real" subject or object, independent of whether the verb happens to occur in an active or a passive sentence.)

Obligatory adverbial complements required by German and English verbs were coded under the subscript RA ("Required Adverb"), whose values specify one or more classes of adverbs, depending on the requirement of the classified verb stem. (These adverb classes are listed on p. III.27.) Optional adverbial complements were indicated only in the case of directional adverbs since this class of adverbs is restricted to co-occurrence with a fairly limited set of verbs.

There is, however, one important difference between verb and adjective valences. Whereas the valence of adjectives is usually restricted to two complements, a large number of verbs has three valences. As indicated, the description of subject and object complements was given under different subscripts, FS and TS for subject and FO and TO for object complements. Features describing multiple object complements were coded in a specially designed format, the descriptors referring to the "direct" object being separated from those of the "indirect" object by the symbol ".". In this respect, the object subscripts FO, TO, and IO (the latter being described below) are different from all others and are referred to as "indexed subscripts". In this format, the descriptors coded for the first object complement under the subscript FO are automatically associated by the system with the descriptors coded for the same complement under the subscripts TO and IO. The same is true of the second complement and any additional ones if more than two object complements should be relevant, as for example in *sich mit jemandem ueber etwas unterhalten*. No special effort was made to decide on which of several complements is the "direct" and which the "indirect" object. For our purposes of automatic analysis and translation, this distinction does not seem to be important. To assure the trans-

lation of multiple objects in their proper relation to the verb it is only necessary to code equivalent objects with identical IO values in the input and output languages. Thus, coding the German verb *bezeichnen* with the features

FO(A.ALS)
TO(E.E)
IO(O.O2)

and its English translation *call* with the features

FO(O.O)
TO(E.E)
IO(O.O2)

guarantees translation of

Sie bezeichnen diesen Prozess als perpetuum mobile

into

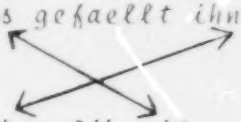
They call this process perpetuum mobile

and not

They call perpetuum mobile this process.

The subscripts IO ("Interpretation of Object") and IS ("Interpretation of Subject") are coded with any of the following values: O ("First Object"), O2 ("Second Object"), and S ("Deep Subject"). For most verbs, these subscripts simply define the described item as first object, second object, or subject of the classified verb and serve mainly to preserve the distinction between the object complements if there are more than one. There are some verbs, however, whose deep subject occurs in the surface sentence as object complement. Usually, this surface object must be rendered as surface subject in another language, as e.g., in

Es gefaellt ihnen.



They like it.

The features

IS(O)
IO(S)

coded with the German verb *gefallen* mark this phenomenon explicitly and thus enable the system to make the proper transformation in the translation process.

Optional object complements are coded with the symbols ",LA", which stand for "or lambda, i.e., nothing". For example, the English verb *employ* is coded with the features

FO(O.AS,LA)
TO(HU.HU,LA)
IO(O.O2,LA)

to indicate the optionality of the *as*-phrase:

The firm employed him as a researcher.

The firm employed him.

Reflexive verbs have the value O4 under IO. This feature signifies the object is to be deleted and thus not to be translated.

An additional subscript TY ("Type") marks those German and English verbs which may not occur in certain constructions. The value NP ("No Passive") identifies verbs which may not be passivized. This feature is not coded for intransitive verbs and verbs which take only reflexive objects since the information would be redundant. The value NG is coded under the subscript "Type" for those English verbs which do not form the progressive.

Like nouns and adjectives, all verbs were coded with inflectional paradigmatic information to guarantee correct analysis and generation of all verb forms. German verbs which form the perfect tense with the auxiliary *sein* were given the feature AX(S); those which take either *sein* or *haben* have the information AX(S,H). The morphological grammar rules for German introduce AX(H) for all verbs which have no subscript AX in the lexicon.

Since the participial forms of English verbs can function as adjectives, they were marked for vocalic or consonantal onset for the correct choice of indefinite article variants.

The subscript PX ("Prefix") contains as values all separable prefixes with which a particular German verb stem may be combined. English verb entries list under the same subscript all permissible adpreps which combine with the specific verb stem. The symbol LA represents the zero adprep or prefix.

Since each stem-adprep or stem-prefix combination forms a new verb with separate features, meaning, and translation equivalents, a feature matrix is associated with every verb. Columns are indicated by " ' ", lines by "/". For example, the English verb *bring* was coded with the features

```
CL(19)
ON(C)
PX(LA    'LA 'ABOUT)/
FS(N     'N  'N,CL)/
TS(HU    'E  'IN,AB,TH)/
FO(O.TO  'O  'O,CL)/
TO(IN.AB 'E  'AB,TH)/
IO(O.O2  'O  'O)/
OA(LA    'DIR'LA)
IS(S)
```

where each vertical column contains the features of a particular verb-adprep combination or reading. By the same method, value sets are divided into separate columns in all dictionary rules where this was necessary to prevent incorrect feature combinations. Separate columns need not represent different readings of a word.

3.2.4 Adverb Classification

In the classification of the German adverbs occurring in our text and of their translation equivalents, only those features were assigned to their dictionary entries which were recognized as necessary for the analysis and translation of the given text.

The semantic type was identified as locative or time by the features TYP(L+ST) (e.g., *here, where*) and TYP(T) (e.g., *today, heute*), respectively. Other features were assigned to identify the lexical categories which each adverb may modify. The values used for this purpose are NUM ("Number", assigned to adverbs like the German *noch*), P ("Positive Adjective Forms", assigned to *so, sehr, very*, etc.), CP ("Comparative Adjective Forms", e.g., *noch*), ADV ("Adverbial, e.g., *erst, also*), N ("Noun Phrase", e.g., *only, auch*), PN ("Noun Phrase, Post-Posed", e.g., *erst, even*), and PAV ("Adverbial, Post-Posed", e.g., *erst, noch*). These tags are referred to in the analysis of a sentence to determine the relationship between the adverb and other sentence constituents, as far as this is possible.

Another important feature, which was assigned only to English adverbs to guarantee correct word order in the output sentence, indicates the position within the sentence in which the classified item has to occur. The positions indicated are pre-verbal, post-verbal, sentence-initial, and sentence-final. Many adverbs were coded with several of these features because they can occur in more than one position within the sentence.

In addition, the class of negative adverbs was identified, as well as the one-member classes of the German item *moeglichst*, in English *as . . . as possible* and the German and English item *so*,

which occurs discontinuously with *that*-clauses and in German in the equative construction *so-wie*. Since English adverbs may occur immediately following the indefinite article, they were also marked for vocalic vs. consonantal onset for the correct choice of *a* and *an*.

3.3 Compounds, Phrases, and Other Problems

3.3.1 Noun Compounds

As was mentioned under "Methodology" on page III.1, LRC lexicographers compiled lists of the German noun compounds (and their English translation equivalents) for the text Raketenantriebe, containing a total of approximately 2,000 compounds.

List I (approx. 1,100 items) contains those compounds which can be analyzed and translated from their components, e.g., *Acetylenherstellung - acetylene production*. It includes only German compounds of the forms $N_1 + N_2$ and A + N; no attempt was made for the time being to analyze verb + noun compounds.

For the compounds on this list, no special dictionary or normal form rules were coded. Instead, the individual components were entered into the dictionary; in Word Analysis they were combined by means of morphological rules, which assign the features of the last component to the resulting compound. This procedure was also applied to German hyphenated compounds and to compounds containing conjunctions, such as *Arbeits- oder Energiegehalt -- work or energy content*.

The great majority of this type of compound can be correctly translated into English compounds of the structure N_1N_2; the example above is of this type. (Compounds which must be rendered as N_2 *of* N_1 in English were not analyzed morphologically but rather treated as members of List II, since the criteria were not established for automatically determining the pattern according to which a compound was to be translated. Example: *Ruhezustand - state of rest.*)

This analysis made it unnecessary to code a separate dictionary rule for each compound, a great advantage in terms of economy. Note that in the text we dealt with, there occurred approximately 2,000 noun compounds and approximately 2,000 simple nouns. The savings will be even greater once morphological rules for other types of compounds have been designed.

List II (approx. 900 items) contains all those German noun compounds which cannot be translated component by component, such as *Serienherstellung - mass production* (not "*series production*"), and those for whose automatic analysis and translation further

research is necessary, e.g., the types mentioned earlier; verb + noun compounds (*Siedepunkt - boiling point*), N_2 *of* N_1 compounds (*Stammland - country of origin*) and others. Since we were not able to analyze these compounds, they were entered as one-word entries, raising the number of noun entries to approx. 3,000. This list will be extremely useful in future research on additional types of compounds.

English compounds which may occur with internal plural inflection, such as *angle(s) of incidence*, were coded as so-called "word-rules". These rules concatenate the sequence *angle* + dummy inflection + *of* + *incidence*, thus accounting for singular and plural forms. (Cf. Appendix II for an explanation of word rules and dummy endings.)

3.3.2 Adjective Compounds

Although adjective compounds occur less frequently than noun compounds, their analysis by morphological compounding rules is nevertheless important for two reasons:

a. Some adjective compounding rules are productive and the number of possible compounds is, therefore, extremely great.

b. The translation of some German adjective compounds results in English adjectival phrases in which the adjective inflection occurs as an internal variable. An example of such a German adjective compound is *energiereich*, whose English translation is *rich in energy*. In the comparative and superlative, the English adjective phrase shows internal inflection: *richer in energy, richest in energy*. It would be impractical to enter all such compounds in all their forms of comparison into the dictionary as separate one-word entries. Therefore, only their components were coded as dictionary entries - in the above case, the noun stems *Energie* and *energ(y)* and the adjective stems *reich* and *rich*. German adjectives which enter into such constructions are marked by the feature FO(NC). The English adjective *rich* (and all others of the same type) was given the feature FO(IN) (cf. 3.2.2 above) to indicate that it takes a complement introduced by the preposition *in*. These stems are concatenated with their proper inflectional endings by the regular adjectival morphology rules of the German and English "word grammars". The English prepositional phrase is constructed as usual and is then concatenated with the preceding, fully inflected adjective. In German, a morphological word rule concatenates the noun with the inflected adjective form, re-writing the string as an adjective which is marked for no further complementation since constructions like **energiereich in* ... are not well-formed.

Other compounds were coded as one-word dictionary entries, e.g., *schwerbelasted (heavily loaded), weitverbreitet (widely used)*. As with noun compounds, the list of these adjective com-

pounds will be used in future work on the internal analysis and translation of adjective compounds.

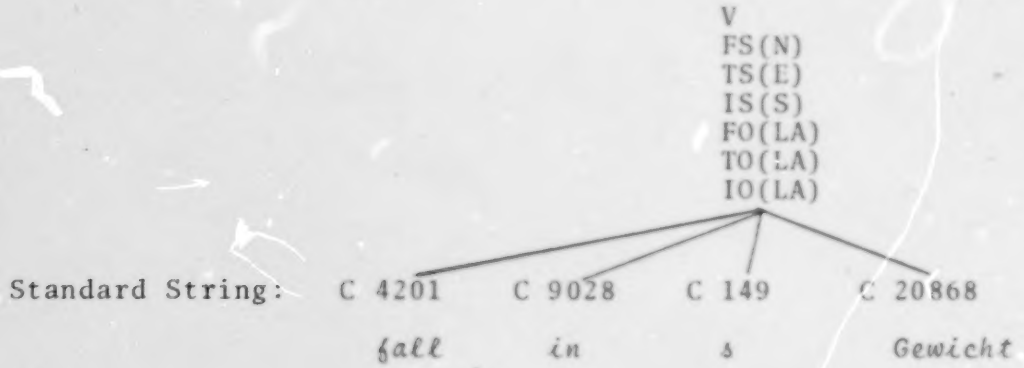
3.3.3 Verbal Lexical Collocations

The problem of verbal lexical collocations and its solution within the LRC system was described in RADC-TR-73-309 Final Technical Report, October 1973. For the reader who is not familiar with this solution, its description is repeated here.

We refer to those verbs as "verbal lexical collocations" which consist of a verb and a noun phrase, a prepositional phrase, a non-finite verb form, an adjective, or an adverb, e.g., *erfolgen = take place; ins Gewicht fallen = be important.* Since such lexical collcations have features and meanings which cannot be derived from their individual components, they must be treated as lexical phrases. However, since their components frequently occur discontinuously and in various permutations, they must be handled differently from normal verbs. For this purpose, the subscript LC ("lexical collocation") was added to the dictionary entries of those nouns and verbs which may occur as components of lexical collocations. Surface analysis refers to these subscripts and guarantees that the items occur contiguously and in a predefined order in the so-called "standard string" (cf. Section II), which is generated after surface analysis.

a. Entries Without Internal Variables

For the actual analysis and translation of lexical collocations, standard dictionary rules were coded which are applied to the standard strings. The constituents of standard strings are the dictionary readings of the underlying lexical elements. The standard dictionary rules concatenate these readings in multibranch rules and assign to the whole structure the syntactic and semo-syntactic features described for the general verb system (c.f. 3.2.3 above). Thus, the German standard dictionary rule C 5011 (marked in Fig. III.9) analyzes the lexical collocation *ins Gewicht fallen (be important)* as follows:



As can be seen, the verb label which dominates the whole string contains features which mark the verb phrase as having only one valence which may be filled by any semantic type of noun: TS(E).

To this standard rule (C 5011) the following German normal form rule applies:

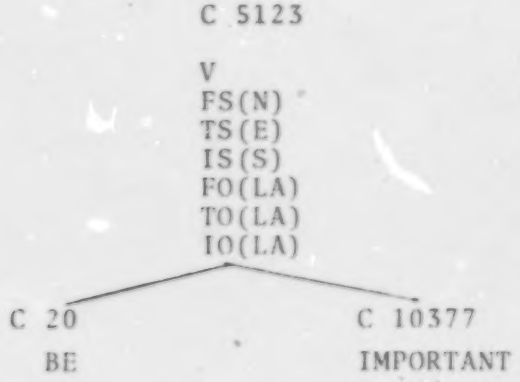
V BE+IMPOR TANT	C 5011	C 20868	C 149 A 1 B 2	C 9028 A 1 B 3	C 4201 A 1 B 4
A CAT(V+P)					
N TM(INS+G EWICHT+F ALLEN)					

[For Fig. III.9 refer to the following page]

The corresponding English normal form rule is in the same translation equivalence class, BE+IMPORTANT:

V BE+IMPOR TANT	C 5173	C 20	C 10377 A 1 B 3
A CAT(V+P)			

The right-hand side of this rule refers to the English standard dictionary rule C 5123, which, in turn, generates the English standard string:



The correct endings and morphological variants (in this case *am*, *are*, *is*, etc.) are generated by the English output syntactic choice algorithm.

German verb phrases which we call "hidden passive phrases" — i.e., those which contain empty function verbs such as *gelangen zu* and *kommen zu*, followed by nominalized verbs — were also treated as lexical collocations. Examples are:

Fig. III.9


```
C 5004    V V          C 4403      Ç 10801
          • TS(MU)     $ PX(DAR)
          • IS(S)
          • FO(A)
          • TO(Ê)
          • IO(U)
          • FS(N)

C 5009    V V          C 4187      Ç 164       C 21619
          • TS(MU)
          • IS(S)
          • FO(AUF1)
          • TO(IN,AB
            ,NT)
          • IO(U)
          • FS(N)

C 5010    V V          C 4201      Ç 9010      C 23170
          • TS(E)      $ PX(LA)
          • IS(S)
          • FO(LA)
          • TO(LÀ)
          • FS(N)
          • IO(LA)

C 5011    V V          C 4201      Ç 9028      C 149       C 20868
          • TS(E)      $ PX(LA)
          • IS(S)
          • FO(LA)
          • TO(LÀ)
          • FS(N)
          • IO(LA)

C 5012    V V          C 4206      Ç 22551
          • TS(MU)     $ PX(LA)
          • IS(S)
          • FO(A)
          • TO(Ê)
          • IO(U)
          • P(LA)
          • FS(N)

C 5013    V V          C 4206      Ç 20000
          • TS(MU)     $ PX(LA)
          • IS(S)
          • FO(A)
          • TO(IN,AB
            ,NT)
          • IO(U)
          • P(LA)
          • FS(N)
```

zur Ausstossung gelangen = *be ejected*

zum Einsatz kommen = *be employed*

An additional subscript P identifies these German phrases as passive in meaning to guarantee their correct translation. The English translation equivalents were not coded as phrases.

b. Lexical Collocations with Internal Variables

Some lexical collocations contain internal variable slots, as for example the noun modifier slot in *to take* V *care that ...*, where V stands for "variable": *he took care that ..., he took great care that ..., he took the greatest possible care that ...*, etc. For such phrases, standard rules were written which provide for variables in their right-hand sides (cf. rule C 5120 below). Since the present rule format does not allow optional rule constituents, several rules were coded for such phrases, one for each possible string. As an example, the English standard rules for *take care that* are shown here:

```
C 5119    V V              C 4726          C 23195
          + FS(N)          $ PX(LA)
          + TS(HU)
          + IS(S)
          + FO(CL)
          + TO(TH)
          + IO(O)

C 5120    V V              C 4726     V ADJ     C 23195
          + FS(N)          $ PX(LA)
          + TS(HU)
          + IS(S)
          + FO(CL)
          + TO(TH)
          + IO(O)
```

where C 4726 is the rule number for *tak(e)* and its stem variants and C 23195 the number for *care*.

For some phrases, up to six rules were necessary to allow for optional determiners, noun modifiers, and plural noun endings, e.g., *to pose* (DET)(ADJ) *problem(s)*. These lexical collocations do not constitute set phrases, but rather, instances in which a verb has a specific and unusual meaning (and translation) in the environment of a noun phrase whose head noun is a particular lexical item. Beyond this, all normal rules of NP analysis and generation apply. A modification of the lexical collocation phase is planned to permit the expression of these relations in a more economical manner than the one described above (cf. Section II).

For the text Raketenantriebe, the number of lexical collocation rules coded is approximately as follows:

German standard dictionary rules:	165
English standard dictionary rules:	160
English normal form rules:	140
German normal form rules:	200

3.3.4 Multi-word Prepositions

Simple multi-word prepositions like the English *with respect to* or the German *auf Grund* (with genitive government) were entered into the English and German dictionaries as unchangeable one-word entries (though with internal blanks).

Multi-word prepositions which occur discontinuously, flanking their objects on both sides, however, cannot be treated in this way by a concatenation system. Examples are in German *nach-hin* (*toward*) and in English *from-on* (*von-ab* or *von-an*). For the analysis of prepositional phrases involving this type of preposition, the simple prepositions like *from* and *nach* were coded with a subscript indicating the specific particles with which they may be combined to form multi-word prepositions.

```
V PREP      = * FROM
+ PR(FROM)
+ AD(ON)

V PREP      = * NACH
+ PR(NACH)
+ PX(HIN,LA)
+ GC(D)
+ TO(E)
+ TYP(PO,PR)
+ ON(C)
```

The particles which combine with them (*on*, *hin*, etc.) were already marked in the dictionary as adpreps (in English) and separable verbal prefixes (in German) with a subscript identifying the lexical item:

```
V PRFX      = * HIN
+ PX(HIN)

V ADPR      = * ON
+ PX(ON)
```

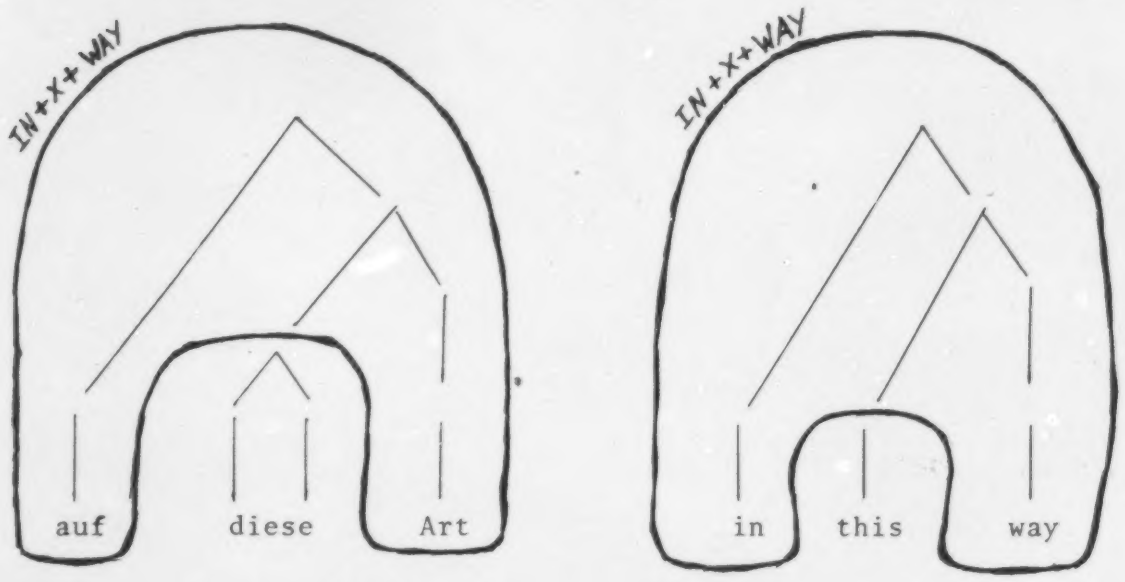
A special grammar rule concatenates the preposition with the following noun phrase (or adverb) and the particle. This rule specifies that the preposition and the particle must have a value of their subscripts PX in common, which guarantees that only proper

combinations are analyzed and generated.

3.3.5 Multi-word Adverbs

As with prepositions, simple multi-word adverbs present no special problem. They were entered into the dictionary as "one-word" entries, e.g., *in addition, on the average, mit einem Schlage.*

Adverbial phrases with internal variable slots were analyzed and generated like regular prepositional phrases by syntactic rules. In order to represent the specific meaning of such phrases, special normal form rules were coded covering the prepositional phrase with its constant items, leaving the variable uncovered. An example of such a phrase is *auf* (determiner and adjective slot) *Art(en)* - *in* (determiner and adjective slot) *way(s)*. The analysis can be represented by the following tree:



3.3.6 Other Problems

The translation problems represented by the list of verbs which must be translated from active into passive or vice versa, and the list of translation of items into different parts of speech, were solved in the syntactic choice and normal form components. Descriptions of these problems and their solutions in the LRC grammar are forthcoming.

Many linguistic phenomena relevant to the classification of dictionary items and automatic translation remain to be researched. One such problem arose in the semantic classification of

nouns and the description of selection restrictions of adjectives. It appears that a number of nouns, which seem to form a natural class (or possibly several classes), may occur with adjectives which do not modify or describe them but rather function as attributes or predicates of other nouns in the same sentence. For example, in the phrase

Holzsprit in gebundener Form

(in English *wood alcohol in bound form*) the adjective *gebunden* does not modify *Form* but rather *Holzsprit*. Similarly, the phrase

das Metall in reinem Zustand

(in English *the metal in its pure state*) can be paraphrased

das reine Metall (the pure metal)

das Metall ist rein (the metal is pure)

but not * *der Zustand ist rein (the state is pure)*

For this reason, the selection restrictions of the adjectives are not satisfied by the semantic type of these nouns. A further example is

Die Substanz hat korrodierende Eigenschaften

(in English: *The substance has corrosive properties*), where the adjective *corrosive* requires a modificand of the type "inanimate physical object", but the noun *property* is of the type "abstract". The real modificand, in this case *substance*, does satisfy the selection restrictions of *corrosive*.

The description of this phenomenon will require further research; the examples collected during the lexicographic work will serve as a basis for such studies.

3.4 Concluding Remarks

The feature system used in the classification of the lexicographic data in the text Raketenantriebe proved sufficient to provide for the disambiguation of a great number of ambiguous words and for the choice of their correct translation equivalents. Unresolved ambiguities will be investigated. We will report on the results of our investigation in future publications.

CONCLUSION

Work completed under this contract has shown the system's capability of providing good mechanical translation of individual sentences. It has shown, at the same time, the limitations of sentence-by-sentence translation and ways how these limitations can be overcome. The experience we gained during the experiment will be used for the modification of algorithms and grammars to obtain operational capability.

The envisioned modifications in algorithms and grammatical descriptions will be made from the point of view of the three criteria for the evaluation of MT: quality, speed and cost. These modifications will not affect the basic design strategy of LRS: machine independence with respect to programming languges, independence of the algorithms from the grammatical description (which permits the incorporation of additional source or target languages without requiring a change of the algorithms); independence of grammatical descriptions from considerations of translatability into a particular target language (which guarantees that any language in LRS can be translated into any other language of LRS, i.e., the bi-directionality of translation).

The modifications in the grammatical description will mainly incorporate the co-occurrence relations between verbs and adverbials to permit LRS to produce even better sentence-by-sentence translations than shown in Appendix I. The quality of the output will be further increased by incorporating a contextual component into LRS which will be able to evaluate the disambiguating information provided in sentence-external context. The speed and thereby the cost of the system's performance will be decreased to a level of economic feasibility by reducing the number of required input and output operations and by omitting a number of translation phases when appropriate. The first objective can be obtained by compacting the lexical data base, by separating the lexicon into a high frequency and normal component, by reducing the amount of information to be stored in the working area and by reducing the number of parasitic readings. The second objective can be obtained by making some of the LRS translation components optional. They will thus only be used when the text actually requires their application.

We expect an operational capability in two-and-a-half to three years with the ability to translate random texts in the sciences. This will be possible due to the availability of a lexical data base for German and English of more than one million scientific and technical terms, supplied by the translation service of the German Federal Bureau of Languages. This data base will be incorporated into the Center's existing lexical data bases of more than 100,000 words for German and English each.

LRS will, in addition, also be sensitive to user specifications by providing translations based on the result of intermediate

analyses with a concomitant reduction in processing time and cost. We will, moreover, consider the possibility of running the System in a time-sharing, multi-processing environment.

The changes and developments of LRS towards operational capability will be described in detail in future reports.

Bibliography

1. Lehmann, Winfred P., and Rolf A. Stachowitz. December 1971. *Feasibility Study on Fully Automatic High Quality Translation.* 2 vols. (RADC-TR-71-295).

2. Stachowitz, Rolf A. "Requirements for Machine Translation: Problems, Solutions, Prospects." 409-532 in Lehmann and Stachowitz, 1971.

3. Lehmann, Winfred P., and Rolf A. Stachowitz. August 1970. *Research in German-English Machine Translation on Syntactic Level, Vol. II.* (RADC-TR-69-368).

4. Lehmann, Winfred P., and Rolf A. Stachowitz. March 1972. *Development of German-English Machine Translation System.* (RADC-TR-72-47).

5. Lehmann, Winfred P., and Rolf A. Stachowitz. April 1972. *Development of German-English Machine Translation System.* (RADC-TR-72-74).

6. Lehmann, Winfred P., and Rolf A. Stachowitz. August 1973. *Development of German-English Machine Translation System.* (RADC-TR-72-260).

7. Lehmann, Winfred P., and Rolf A. Stachowitz. October 1973. *Development of German-English Machine Translation System.* (RADC-TR-73-309).

8. Stemmer, Josef. 1952. *Raketenantriebe: Ihre Entwicklung, Anwendung and Zukunft.* Zurich: Schweizer Druck- und Verlagshaus. 523 pp.

9. Stachowitz, Rolf A. "Beyond the Feasibility Study: Syntax and Semantics." Paper presented at the July 1972 Meeting of the Assn. for Computational Linguistics. Austin: LRC, 1972. 14 pp.

10. Stachowitz, Rolf A. "Lexical Features in Translation and Paraphrasing: An Experiment." 367-408 in Lehmann and Stachowitz, 1971.

APPENDIX I

For the demonstration, 42 sentences (three pages) of a 150-page excerpt of a text on rocket propulsion were translated. The dictionary was prepared for the whole excerpt. For each word, all translations appropriate in the text were listed; in addition, the usual dictionary translations were added to nouns and adjectives to show the system's capability for disambiguation. No additional dictionary translations were listed for verbs, since they had a large number of translation equivalents from the start because of the frequent occurrence of German verb/prefix combinations.

The German surface syntax was restricted to the description of the structures in the three pages. The syntactic choice grammar (reverse transformations) covers the greater part of the German grammar. The standard (deep) grammars for German and English were complete.

The output format is a simulation of the computer output. Only a few sentences were translated in this format since the contract terminated before the completion of the debugging of the program "syntactic choice output" (forward transformations) and of the print-out program (lexical spell-out) which effects the columnization of translation alternatives. All sentences were, however, mechanically translated into "standard" English (the English deep representation).

S 1

2 $DIE $ENERGIETRAEGER
2 $THE ENERGY CARRIERS

S 2

$DAS $ANTRIEBSGESETZ DES $RAKETENMOTORS.
$THE DRIVE PROPULSION LAW OF THE ROCKET MOTOR.

S 3

$ALS $RUECKSTOSSMOTOR KANN JEDE $MASCHINE BEZEICHNET WERDEN, BEI WELCHER EIN $GAS ODER EINE $FLUESSIGKEIT UNTER $DRUCK MIT HOHER $GESCHWINDIGKEIT AUS EINER $OEFFNUNG AUSSTROEMT UND DADURCH EINE $REAKTIONS- ODER $RUECKSTOSSKRAFT HERVORRUFT.
$ANY EACH EVERY MACHINE MACHINERY WHERE A GAS OR A LIQUID FLOWS OUT FROM AN OPENING UNDER PRESSURE WITH HIGH VELOCITY AND THEREBY CAUSES A REACTION OR RECOIL FORCE CAN BE CALLED A JET PROPULSION MOTOR.
POWER

S 4

$ZUR $VERANSCHAULICHUNG DIESES $VORGANGES DIENE FOLGENDES $BEISPIEL.
$THE FOLLOWING EXAMPLE MAY SERVE FOR THE ILLUSTRATION OF THIS PROCESS.

S 5

$IN EINEM ALLSEITIG GESCHLOSSENEN $ZYLINDER BEFINDET SICH EIN $GAS UNTER $DRUCK.
$A GAS IS UNDER PRESSURE IN A CYLINDER CLOSED ON ALL SIDES.

S 6

$DAS $GAS WIRD SICH IN DEM $BEHAELTER GLEICHMAESSIG VERTEILEN UND DER $DRUCK IST AN ALLEN $TEILEN DER $INNENFLAECHE GLEICH GROSS.

$THE GAS WILL SPREAD / DISTRIBUTE IN THE CONTAINER UNIFORMLY AND THE PRESSURE IS EQUALLY GREAT / HIGH / EXTENSIVE AT ALL PARTS OF THE INNER SURFACE.

S7

$DIE $DRUCKWIRKUNG LAESST SICH SO DEFINIEREN, DASS DAS $GAS AUS EINZELNEN $MOLEKUELEN BESTEHT, WELCHE IN $WIRKLICHKEIT SO KLEIN SIND, DASS WIR SIE MIT BLOSSEM $AUGE NICHT SEHEN KOENNEN.

$THE PRESSURE EFFECT CAN BE DEFINED IN SUCH A WAY THAT THE GAS CONSISTS OF INDIVIDUAL MOLECULES WHICH ARE SO SMALL IN REALITY THAT WE CANNOT SEE THEM WITH THE NAKED EYE.

S 8

$WIR WOLLEN SIE UNS ABER ALS $KUGELN VORSTELLEN.

$BUT WE WANT TO IMAGINE THEM AS BALLS / GLOBES / SPHERES.

S 9

$ZWISCHEN DIESEN EINZELNEN $MOLEKUELEN ODER $KUGELN HERRSCHT EINE $SPANNUNG ODER $DRUCK, WELCHEN WIR BILDLICH MIT EINER $SPIRALFEDER DARSTELLEN KOENNEN.

$A TENSION OR PRESSURE WHICH WE CAN REPRESENT FIGURATIVELY WITH A SPIRAL SPRING PREVAILS AMONG THESE INDIVIDUAL MOLECULES OR BALLS / GLOBES / SPHERES.

S 10

$ENTSPRECHEND EINER GROSSEN $ANZAHL VON $MOLEKUELEN BEFINDET SICH IM $ZYLINDER EINE NOCH GROESSERE $ANZAHL VON $SPIRALFEDERN.

$AN EVEN GREATER / HIGHER / MORE EXTENSIVE NUMBER OF SPIRAL SPRINGS IS CORRESPONDING TO A GREAT / HIGH / EXTENSIVE NUMBER OF MOLECULES IN THE CYLINDER.

S 11

$ABB. 32. $DRUCKWIRKUNG IN EINEM $BEHAELTER.

$FIGURE / ILLUSTRATION 32. $PRESSURE EFFECT / RESULT / ACTION IN A CONTAINER.

S 12

$DER $BEHAELTER ENTHAELT DEMZUFOLGE GROSSE $KRAEFTE, WELCHE GLEICHMAESSIG AUF DIE $ZYLINDERWANDUNGEN WIRKEN.

$THE CONTAINER CONSEQUENTLY CONTAINS GREAT / HIGH / EXTENSIVE FORCES / POWERS WHICH AFFECT / ACT / HAVE AN EFFECT ON THE CYLINDER WALLS UNIFORMLY.

S 13

$DIE $KRAEFTE WIRKEN NATUERLICH AUCH DANN, WENN SICH KEINE $LUFT IN DER $UMGEBUNG DES $ROHRES BEFINDET.

$THE FORCES / POWERS NATURALLY HAVE AN EFFECT EVEN WHEN NO AIR EXISTS IN THE ENVIRONMENT OF THE TUBE / BARREL / PIPE

S 14

$SOLANGE DIESES GESCHLOSSEN BLEIBT, BESTEHT EIN $RUHEZUSTAND.

$A STATE OF REST EXISTS AS LONG AS THIS REMAINS CLOSED.

S 15

$ERST WENN WIR EINE $ZYLINDERWAND WEGNEHMEN, STROEMEN DIE EINZELNEN $GASTEILCHEN INFOLGE DER ZWISCHEN IHNEN WIRKENDEN $SPANNKRAFT DES $DRUCKES MIT GROSSER $GESCHWINDIGKEIT AUS DEM $ROHRE.

$THE INDIVIDUAL GAS PARTICLES FLOW OUT FROM THE TUBE / PIPE / BARREL WITH GREAT / HIGH / EXTENSIVE VELOCITY BECAUSE OF THE TENSION OF THE PRESSURE HAVING AN EFFECT AMONG THEM ONLY WHEN WE TAKE OFF / AWAY A CYLINDER WALL.

S 16

$DAMIT TRITT DAS FUNDAMENTALE PHYSIKALISCHE $GESETZ <<$ACTIO ET $REACTIO>> IN $KRAFT.

$THE FUNDAMENTAL PHYSICAL LAW <<ACTIO ET REACTIO>> THEREBY BECOMES EFFECTIVE.

S 17

$DAS HEISST, JEDE $WIRKUNG HAT EINE GLEICH GROSSE, ABER ENTGEGENGESETZT GERICHTETE $GEGENWIRKUNG ZUR $FOLGE.

$THAT IS, ANY / EVERY / EACH EFFECT / RESULT / ACTION CAUSES AN EQUALLY HIGH / GREAT / EXTENSIVE , BUT OPPOSITELY DIRECTED REACTION / COUNTEREFFECT

S 18

$DIESES $GESETZ GILT AUCH FUER DEN LUFTLEEREN $RAUM, WAS FUER UNS PRAKTISCH BEDEUTET, DASS WIR MIT EINEM $RUECKSTOSSMOTOR IN UNBEGRENZTE $HOEHEN, IN DIE $STRATOSPHAERE ODER GAR IN DEN $WELTENRAUM FLIEGEN KOENNEN.

$THIS LAW HOLDS ALSO FOR THE EMPTY ROOM / SPACE / CABIN / COMPARTMENT , WHICH PRACTICALLY MEANS FOR US THAT WE CAN FLY WITH A JET PROPULSION MOTOR INTO UNLIMITED HEIGHTS / ALTITUDES , INTO THE STRATOSPHERE OR EVEN INTO THE OUTER SPACE.

S 19

$DENN DAS $GAS STROEMT AUS DER $ZYLINDEROEFFNUNG HERAUS, AUCH WENN KEINE UMGEBENDE $LUFT DA IST.

$FOR THE GAS STREAMS OUT OF THE CYLINDER OPENING EVEN WHEN NO SURROUNDING AIR EXISTS.

S 20

$DIE $GASMOLEKUELE STOSSEN SICH NICHT AN DIESER AB, WIE HEUTE NOCH VIELFACH ARGUMENTIERT WIRD.

$THE GAS MOLECULES DO NOT PUSH OFF FROM THIS AS IS ARGUED FREQUENTLY STILL TODAY.

S 21.1

$DIE RUECKSTOSSKRAFT KOMMT EINFACH ZUSTANDE, INDEM SICH JEDE $SPANNKRAFT ($FEDER) ZWISCHEN DEN $MOLEKUELEN ($KUGELN) NACH DER $ZYLINDEROEFFNUNG HIN ENTSPANNT.

$THE RECOIL FORCE COMES ABOUT SIMPLY AS ANY / POWER / EACH / EVERY TENSION (SPRING) RELAXES AMONG THE MOLECULES (BALLS) / (GLOBES) / (SPHERES) TOWARD THE CYLINDER OPENING.

S 21.2

$THE RECOIL FORCE COMES ABOUT SIMPLY AS ANY / POWER / EACH / EVERY TENSION (SPRING) AMONG THE MOLECULES (BALLS) / (GLOBES) / (SPHERES) RELAXES TOWARD THE CYLINDER OPENING.

S 22

$GEGEN DIE VORDERE, GESCHLOSSENE $STIRNWAND WIRKT DER VORHANDENE $GASDRUCK AUF DIE GESAMTE $FLAECHE, NACH DER $OEFFNUNG HIN ABER NUR AUF DIE UM DIESEN $QUERSCHNITT VERKLEINERTE $GRUNDFLAECHE.

$THE PRESENT / AVAILABLE GAS PRESSURE AFFECTS / ACTS / HAS AN EFFECT ON THE ENTIRE SURFACE TOWARD THE ANTERIOR, CLOSED FRONT WALL BUT ONLY ON THE SURFACE AREA REDUCED BY THIS CROSS SECTION TOWARD THE OPENING.

S 23

$DIESER $DRUCKUNTERSCHIED, HERVORGERUFEN DURCH DIE $EXPANSION, BEWIRKT SOMIT AUCH, DASS DAS $POHR NACH DER DER $AUSSTROEMRICHTUNG ENTGEGENGESETZTEN $SEITE WEGFLIEGT.

$THIS PRESSURE DIFFERENCE CAUSED BY THE EXPANSION THEREFORE CAUSES ALSO THE TUBE / PIPE / BARREL TO FLY AWAY TOWARD THE SIDE PAGE OPPOSITE TO THE DIRECTION OF EXHAUST.

S 24

$DAS $RUECKSTOSSPRINZIP IST WEITERHIN EIN $SPEZIALFALL DES $SATZES VON DER $ERHALTUNG DES $SCHWERPUNKTES, DER SEINERSEITS WIEDER AUF DAS DRITTE $NEWTONSCHE $GRUNDGESETZ VON $ACTIO ET $REACTIO ZURUECKZUFUEHREN IST.

$THE RECOIL PRINCIPLE FURTHERMORE IS A SPECIAL CASE OF THE CHARGE SENTENCE PRINCIPLE OF THE PRESERVATION OF THE CENTER OF GRAVITY WHICH ON ITS PART CAN BE TRACED BACK AGAIN TO THE THIRD $NEWTONIAN BASIC LAW OF ACTIO ET REACTIO.

S 25

$AUS DER $NEWTONSCHEN $BEWEGUNGSGLEICHUNG LAESST SICH DER $IMPULSSATZ HERLEITEN, WELCHER FUER ALLE $STOSSVORGAENGE $ANWENDUNG FINDET.

$THE IMPULSE SENTENCE PRINCIPLE WHICH APPLIES FOR ALL PERCUSSION PROCESSES CAN BE DERIVED FROM THE $NEWTONIAN KINETIC EQUATION.

S 26.1

$WIE IN DER $FOLGE GEZEIGT WIRD, KANN DIE $GAS- UND $DRUCKERZEUGUNG IM $ZYLINDER AUF VIELE $ARTEN ERFOLGEN.

$THE GAS AND PRESSURE PRODUCTION GENERATION IN THE CYLINDER CAN TAKE PLACE IN MANY WAYS AS IS SHOWN SUBSEQUENTLY.

S 26.2

$THE GAS AND PRESSURE PRODUCTION GENERATION CAN TAKE PLACE IN THE CYLINDER IN MANY WAYS AS IS SHOWN SUBSEQUENTLY.

S 27

$DURCH $VERBRENNUNG EINES BRENNBAREN FESTEN $STOFFES ODER FLUESSIGEN $TREIBSTOFFGEMISCHES IN EINER $BRENNKAMMER IST MAN IN DER $LAGE, DIE NOTWENDIGE $AUSSTROEMMASSE UND $DRUCKDIFFERENZ ZU BESCHAFFEN.

$ONE CAN PROVIDE THE NECESSARY EXHAUST MASS AND PRESSURE DIFFERENCE THROUGH THE COMBUSTION OF A COMBUSTIBLE FIRM SOLID SUBSTANCE OR LIQUID FUEL MIXTURE IN A COMBUSTION CHAMBER.

S 28.1

$HANDELT ES SICH UM $FLUGKOERPER, WELCHE IN UNBEGRENZTE $FLUGHOEHEN STEIGEN SOLLEN, SO IST ES NOTWENDIG, DASS DER ZUR $VERBRENNUNG NOTWENDIGE $SAUERSTOFF IM $FLUGKOERPER SELBER MITGEFUEHRT WIRD.

$IF WE ARE DEALING WITH MISSILES WHICH SHALL ARE TO RISE INTO UNLIMITED FLIGHT HEIGHTS ALTITUDES THEN IT IS NECESSARY THAT THE OXYGEN NECESSARY FOR THE COMBUSTION IN THE MISSILE ITSELF IS CARRIED ALONG.

S 28.2

$IF WE ARE DEALING WITH MISSILES WHICH SHALL ARE TO RISE INTO UNLIMITED FLIGHT HEIGHTS ALTITUDES THEN IT IS NECESSARY THAT THE OXYGEN NECESSARY FOR THE COMBUSTION IS CARRIED ALONG IN THE MISSILE ITSELF.

S 29

$ES HANDELT SICH HIER DANN ALSO UM EIN VON DER $UMGEBUNG VOELLIG UNABHAENGIGES $MOTORSYSTEM.

$WE ARE DEALING HERE THEN WITH A MOTOR SYSTEM COMPLETELY INDEPENDENT OF THE ENVIRONMENT.

S 30

$GELINGT DIE $SCHAFFUNG EINES SOLCHEN, DANN IST AUCH DIE $ANFLIEGUNG ANDERER $HIMMELSKOERPER MOEGLICH.

$IF THE CREATION OF SUCH SUCCEEDS THEN ALSO THE FLIGHT TO DIFFERENT HEAVENLY BODIES IS POSSIBLE.

S 31

$DER VORLIEGENDE ERSTE $BAND UEBER $RAKETENANTRIEBE BEFASST SICH AUSSCHLIESSLICH MIT SELBSTAENDIGEN $MOTORSYSTEMEN, WELCHE GENERELL IN DEN $PULVERRAKETEN, $RAKETENMOTOREN MIT FLUESSIGN $TREIBSTOFFEN SOWIE ATOMAREN UND ELEKTRISCHEN $SYSTEMEN ZUSAMMENGEFASST WERDEN KOENNEN.

$THE PRESENT FIRST VOLUME ON ROCKET DRIVES PROPULSION EXCLUSIVELY DEALS CONCERNS ITSELF WITH INDEPENDENT MOTOR SYSTEMS WHICH GENERALLY CAN BE SUBSUMED UNDER THE DRY FUEL ROCKETS, ROCKET MOTORS WITH LIQUID FUELS AS WELL AS ATOMIC NUCLEAR AND ELECTRICAL SYSTEMS.

S 32

$TURBO- ODER $DUESENANTRIEBE SIND AN EINE $LUFTHUELLE GEBUNDEN, DA DER ZUR $VERBRENNUNG NOTWENDIGE $SAUERSTOFF AUS DIESER ENTNOMMEN WIRD.

$TURBO- OR JET PROPULSION SYSTEMS ARE BOUND TO AN ATMOSPHERE SINCE THE OXYGEN NECESSARY FOR THE COMBUSTION IS TAKEN EXTRACTED FROM IT.

S 33

$UEBER SOLCHE UNSELBSTAENDIGE $SYSTEME WIRD IN EINEM ZWEITEN $BAND BERICHTET.
$SUCH DEPENDENT SYSTEMS ARE REPORTED IN A SECOND VOLUME.

S 34

$WIE ERWAEHNT, GEHOERT AUCH DIE $PULVERRAKETE ZU DEN SELBSTAENDIGEN $RUECKSTOSS-$FLUGSYSTEMEN.
$ALSO THE DRY FUEL ROCKET BELONGS TO THE INDEPENDENT RECOIL FLIGHT SYSTEMS AS HAS BEEN MENTIONED.

S 35

$DER ZUR $VERBRENNUNG NOTWENDIGE $SAUERSTOFF IST IN EINEM $BESTANDTEIL DES $PULVERS, DEM $SALPETERSALZ, ENTHALTEN.
$THE OXYGEN NECESSARY FOR THE COMBUSTION IS CONTAINED IN A COMPONENT OF THE POWDER, THE NITRATE.

S 36

$DIE HEUTE UEBLICHEN $KONSTRUKTIONEN VON $RUECKSTOSSTRIEBWERKEN, WELCHE MIT $PULVERN ODER FLUESSIGEN $TREIBSTOFFEN
BETRIEBEN WERDEN, BENENNT MAN OFT AUCH ALS $GASDRUCK-$RAKETENANTRIEBE.
$ONE OFTEN CALLS THE MODELS / CONSTRUCTIONS OF JET PROPULSION POWER UNITS CUSTOMARY TODAY WHICH ARE OPERATED WITH POWDERS OR
LIQUID FUELS ALSO GAS PRESSURE ROCKET DRIVES.

S 37

$DIE $GROESSE DER ERZIELBAREN $RUECKSTOSSKRAFT HAENGT VON DER $ART DES $STOFFES AB, WELCHER DURCH DIE $OEFFNUNG EINES
$RAKETENROHRES AUSSTROEMT, SOWIE VON DER $FORMGEBUNG DIESER $OEFFNUNG SELBST.
$THE SIZE OF THE ATTAINABLE RECOIL FORCE / POWER DEPENDS ON THE TYPE OF THE SUBSTANCE WHICH FLOWS OUT THROUGH THE OPENING
OF A ROCKET TUBE / PIPE / BARREL AS WELL AS ON THE SHAPE OF THIS OPENING ITSELF.

S 38

$WUERDE MAN LEDIGLICH AUS EINEM $ZYLINDER $DRUCKLUFT ODER $DRUCKWASSER AUS EINER GEWOEHNLICHEN $OFFNUNG AUSSTROEMEN LASSEN, SO IST DIE ERZIELBARE $KRAFT GERING, WEIL DIE $AUSSTROEMGESCHWINDIGKEIT KLEIN IST.

$IF ONE ONLY CAUSED COMPRESSED AIR OR PRESSURIZED WATER TO FLOW OUT FROM THE CYLINDER OUT OF A USUAL OPENING THEN AN ORDINARY

THE ATTAINABLE FORCE WOULD BE LOW POWER SMALL LITTLE BECAUSE THE EXHAUST VELOCITY IS LOW SMALL LITTLE .

S 39

$WIRD HINGEGEN IN DIE $DRUCKLUFT NOCH IRGENDEINE BRENNBARE $FLUESSIGKEIT EINGESPRITZT UND DIESES $GEMISCH ENTZUENDET, SO STROEMEN DIE $VERBRENNUNGSGASE MIT SEHR GROSSER $GESCHWINDIGKEIT AUS, WEIL DURCH DIE EXPLOSIONSARTIGE $VERBRENNUNG INNERHALB DES $ZYLINDERS EIN GROSSER $DRUCKANSTIEG ERREICHT WIRD.

$IF HOWEVER ANY SOME OTHER COMBUSTIBLE LIQUID IS INJECTED INTO THE COMPRESSED AIR AND THIS MIXTURE IS IGNITED THEN THE

COMBUSTION GASES FLOW OUT WITH VERY HIGH GREAT EXTENSIVE VELOCITY BECAUSE A HIGH GREAT AN EXTENSIVE PRESSURE ASCENT INCREASE IS ATTAINED ACHIEVED THROUGH THE

EXPLOSIVE COMBUSTION WITHIN THE CYLINDER.

S 40

$DAS GLEICHE IST AUCH BEI DER $VERBRENNUNG VON $PULVER IN EINER $RAKETENHUELSE DER $FALL.

$THE SAME EQUAL IS THE CASE ALSO DURING THE COMBUSTION OF POWDER IN A ROCKET SHELL.

S 41

$DAS GANZE $PROBLEM ZIELT SOMIT DARAUF HIN, MOEGLICHST KRAEFTIGE, DAS HEISST ENERGIEREICHE $TREIBSTOFFE FUER $RAKETENMOTORE ZU VERWENDEN.

$THE WHOLE PROBLEM THEREFORE POINTS TOWARD USING FUELS AS POWERFUL, I.E. THAT IS AS RICH IN ENERGY AS POSSIBLE FOR ROCKET

MOTORS.

S 42

$AUS DIESEM $GRUNDE SOLLEN VORERST SOLCHE $STOFFE UND DEREN $VERBRENNUNGSEIGENSCHAFTEN NAEHER BETRACHTET WERDEN.
$SUCH SUBSTANCES AND THEIR COMBUSTION PROPERTIES SHALL BE INVESTIGATED FOR THIS REASON FIRST OF ALL MORE CLOSELY.

MISSION
of
Rome Air Development Center

RADC is the principal AFSC organization charged with planning and executing the USAF exploratory and advanced development programs for information sciences, intelligence, command, control and communications technology, products and services oriented to the needs of the USAF. Primary RADC mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, and electronic reliability, maintainability and compatibility. RADC has mission responsibility as assigned by AFSC for demonstration and acquisition of selected subsystems and systems in the intelligence, mapping, charting, command, control and communications areas.